

No Freight Unhoused: Leveraging Data to Identify Underserved Markets

by

Mark Featherstone

BS, Supply Chain and Information Systems, Pennsylvania State University

and

Lydia Whipple

BA, Supply Chain Management, Michigan State University

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

© 2025 Mark Featherstone and Lydia Whipple. All rights reserved.

The authors hereby grant to MIT permission to reproduce and to distribute publicly paper and electronic copies of this capstone document in whole or in part in any medium now known or hereafter created.

Signature of Author: _____
Mark Featherstone
Department of Supply Chain Management
May 9, 2025

Signature of Author: _____
Lydia Whipple
Department of Supply Chain Management
May 9, 2025

Certified by: _____
Dr. Ilya Jackson
Research Scientist
Capstone Advisor

Accepted by: _____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

No Freight Unhoused: Leveraging Data to Identify Underserved Markets

by

Mark Featherstone

and

Lydia Whipple

Submitted to the Program in Supply Chain Management
on May 9, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

This project is in collaboration with a private equity firm that invests in real estate used for transportation and logistics. The goal of this research is to determine underserved geographic markets where the amount of freight activity is greater than the amount of supply chain real estate to support it. Supply chain real estate investments have historically been focused in or near large urban centers, and this project seeks to use a data-driven approach to identify markets that could be advantageous for investment. Our approach consisted of three key steps, which include mapping transportation flows in the United States, understanding the sponsor company's portfolio and overall supply chain real estate market, and analyzing underserved markets by considering both transportation flows and supply chain real estate data. Regions of the United States were aggregated into Key Market Areas (KMAs) and then ranked from most to least underserved based on a ratio of freight activity to supply chain real estate within each KMA. A wide range of ratios were calculated as the most overserved market's ratio is over 40 times that of the least underserved market. The final output is a tool that visualizes underserved geographic markets on a dashboard which can be used to support investment decisions by the firm.

Capstone Advisor: Dr. Ilya Jackson

Title: Research Scientist, MIT Center for Transportation and Logistics

ACKNOWLEDGMENTS

We want to take this opportunity to thank our advisor, Dr. Ilya Jackson, for his time and expertise. We would not have been able to navigate this capstone without his commitment to supporting us and willingness to help. We also would like to thank Dr. Angi Acocella for her time and input. Lastly, we want to give our utmost thanks to our sponsor company for their continued collaboration with MIT and time committed to supporting us on this project. It was a wonderful journey with some amazing insights that we hope continue to help the team in the future!

Table of Contents

1. INTRODUCTION	6
1.1 Background	6
1.1.1 Transportation Modes	6
1.1.2 Logistics Providers	7
1.1.3 Inbound vs. Outbound Logistics	7
1.1.4 High-Flow-Through Properties	8
1.1.5 Supply Chain Real Estate Facility Types	8
1.1.6 Key Market Areas	9
1.2 Motivation	10
1.3 Problem Statement and Key Questions	11
1.4 Goals and Expected Outcomes	11
1.5 Plan of Work	12
2. STATE OF THE PRACTICE	13
2.1 Logistics Trends Impacting Supply Chain Real Estate	13
2.2 Supply Chain Real Estate	13
2.3 Transportation Flows Impacting SCRE	14
3. METHODOLOGY	15
3.1 Collect Data on Transportation Flows	16
3.2 Analyze Transportation Flows Data	17
3.3 Assess Sponsor Company's Portfolio	20
3.4 Build Visualizations	22
3.5 Merge Data Sources and Develop Ratios	25
3.6 Quantify Underserved Markets with Merged Data	26
4. RESULTS	27
4.1 The Interactive Dashboard	27
4.2 Ratio Results	30
4.3 Interpreting Ratios	32
4.4 Contributions to Sponsor	33
4.5 Limitations and Areas of Improvement	34
5. CONCLUSION	35
5.1 Insights and Recommendations	36
5.2 Future Applications	36

1. INTRODUCTION

Freight transportation plays an important role in the economy as it links suppliers with buyers of goods in the US and around the world. In 2022, the trucking industry accounted for nearly \$1 trillion in gross freight revenue (Elgin, 2024). Goods are often transported by multiple modes by the time they reach their end-customer. From sourcing raw materials to final assembly and delivery, various storage and warehousing of goods are needed. Our sponsor company is a private equity firm focused on selecting and purchasing supply chain real estate (SCRE) facilities like truck terminals, secure storage and parking, and final mile delivery facilities, each of which supports a different stage of production and distribution of products. They purchase properties and then lease them to clients, who often operate in logistics and retail industries. Transportation flows and SCRE are interconnected; topics like transportation modes, supply chain operations, and US logistics trends help explain their relationship.

1.1 Background

The following subsections provide an overview of the freight transportation landscape, including types of transportation modes, roles of logistics providers, and nuances between inbound and outbound logistics. Freight transportation helps explain the importance of high-flow-through (HFT) and other SCRE facility types which are discussed later in this section.

1.1.1 Transportation Modes

Various transportation modes are used for inland freight transportation in the US. Truckload, less than truckload (LTL), drayage, and parcel are four transportation modes that impact the sponsor's SCRE portfolio.

Truckload transportation is used when a shipper fills an entire truck with their goods in a single shipment. Since the whole truck is dedicated to one shipper, stops and delays can be minimized, which helps reduce transit time. Truckload shipments are commonly used for large shipments or for time-sensitive goods that need to reach their destination quickly. This mode of transportation has multiple sub-categories, including dry van, reefer, and flatbed (Rodrigue, 2020). Less than truckload, commonly known as LTL, is used by shippers for midsize shipments when they are not able to fill up an entire truck. LTL shipments consolidate multiple partial loads from shippers into one truck, sharing space and expenses. LTL shipments have multiple stops, so a key tradeoff for shippers is a longer lead time for a lower cost (Rodrigue, 2020).

Drayage is used for short-distance transportation of goods after they have arrived at a port. Shippers use this mode when they have goods that enter a port but want to store them at nearby facilities before shipping them to their next destination. Drayage is an important intermediate step between transportation modes like ocean and truckload (Maersk, 2024). It is also an important consideration for our capstone sponsor as they invest in facilities near ports. Parcel transportation is used for small shipments that often reach the destination of the end-customer and has boomed in the e-commerce era (Nahata, 2022). Our sponsor has investments in final mile delivery facilities, so it is a crucial freight transportation mode for them.

1.1.2 Logistics Providers

Shippers typically have a choice between carriers that directly service their freight transportation needs or brokers who do not own any trucks but facilitate shipment transactions between shippers and carriers. In recent years, digital freight brokers (DFBs) have emerged as a new class of non-asset-based providers. DFBs are more technology-based than traditional brokerages and have increased market share in recent years (Acocella, 2022). DFBs serve transportation modes like truckload, LTL, drayage, and parcel.

1.1.3 Inbound vs. Outbound Logistics

At the micro-level of our sponsor's SCRE portfolio and the macro-level of the transportation industry in the US, it is important to understand inbound vs. outbound logistics. Both are important transportation flows within a supply chain network but provide different insights into the manufacturing and consumption activities of a geographic area.

Inbound logistics is the process of moving goods into a geographic area, usually to a distribution center or other supply chain facility. Shipments enter a geographic area for a variety of reasons. Raw materials that will later be assembled into a finished product may arrive. Conversely, finished goods distributed to end-customers through a retail or direct-to-consumer channel may enter the area.

Outbound logistics is the process of moving goods out of a geographic area. These shipments often originate in a manufacturing facility or distribution center and contain raw materials or finished goods. For smaller shipments like parcels, outbound logistics likely indicate that the goods are moving closer and possibly directly to the end-customer (Jenkins, 2023).

1.1.4 High-Flow-Through Properties

Our sponsor aims to identify geographic areas with fruitful real estate investment opportunities and purchase SCRE in those locations. The sponsor then leases their properties out to companies over a long contract period of typically five or more years. Companies can choose to own or lease SCRE properties, but in many cases chooses to lease due to the increased flexibility from limited capital investment (Jacobs, 2021). This is especially true in the case of high-flow-through (HFT) properties, where goods remain on-site for a maximum of 24 hours before moving to the next location.

HFT properties often serve as a transition point in the supply chain. The transition can be between drivers, a way to reposition trailers, or as a location for the final breakdown of a pallet. These properties tend to address the needs of a wide variety of customers and often have a smaller building-to-land ratio than a traditional real estate investment may have (Cauble, 2023).

1.1.5 Supply Chain Real Estate Facility Types

In this project, the sponsor's SCRE facility types are segmented into three categories – truck terminals, secure storage and parking, and final mile delivery facilities.

The sponsor company defines truck terminals as transfer or consolidation points in the movement of goods across the US. These terminals are often located along major highways. They comprise a long, narrow building with an equal number of bay doors on either side, a wide path for trucks to pass each other, and a large amount of parking space for trailers. The building site coverage is typically around 15% of the property and is filled and emptied twice per day with goods that are only temporarily stored, without racks, before transferring to another trailer (Sponsor Company, personal communication, October 2024). A truck terminal can flexibly serve multiple freight transportation modes and customer types simultaneously.

The second SCRE type our sponsor owns is secure storage and parking properties. Often, this type includes large properties with highway access, many fueling points, offices, maintenance facilities, and trailer parking spots. These properties can be used as a transfer point for drayage freight, a repositioning point for empty containers, or a maintenance point for a company's fleet.

Lastly, the sponsor has final mile delivery properties. This type of SCRE is used for parcels, e-commerce, special handling (also referred to as “white glove” service), and food/beverage that

must be sorted and positioned prior to delivery to the customer. Final mile delivery properties generally have a 50–80% site coverage ratio, with the property allowing for trucks/trailers, delivery vans, and employee parking all tied to a warehouse or sorting center that will process the shipments (Sponsor Company, personal communication, October 2024). Products will have high turnover in the building, which will require more customization of an efficient layout, depending on the customers' needs. With the rise in e-commerce, there is higher demand for final mile delivery services today. However, the final mile delivery SCRE has more specific customer requirements tied to location and building setup, which makes it the most complex SCRE type that our sponsor focuses on.

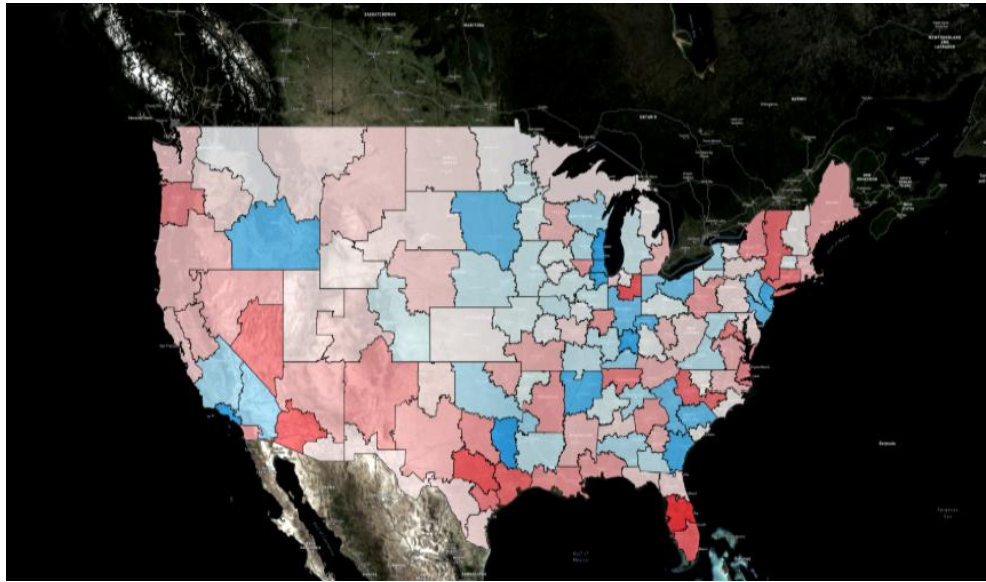
The focus of this project is on truck terminals and secure storage and parking, as those most closely align with longer freight hauls, and transportation flows data is more readily available from both federal and private entities. Conversely, this project is not focused as much on final-mile delivery sites, as their value is more tied to consumer habits in a geographic region, and they require a much closer location to the customer (Prologis, 2019). In this capstone, we sought to understand how transportation flows on a broader level are changing and therefore impact the strategy for investing in SCRE.

1.1.6 Key Market Areas

A key market area (KMA) is a collection of 3-digit ZIP code areas, and each revolves around a major city or economic center. KMA is a standard for many companies in determining freight pricing (Pitz, 2017) and allows aggregation across ZIP codes of five or more digits. There are 135 KMAs in the US.

In this project, transportation flows are analyzed at the KMA-level. This allows for a simpler analysis of the state of the trucking industry, because only 135 KMAs are considered compared to the tens of thousands of ZIP codes in the US. Additionally, there are no state boundaries assigned to a KMA, so they properly represent a region relative to the concentration of trucking movement trends, rather than arbitrary bounds. This is an important level of data granularity in the case of truck terminals and secure storage and parking facilities because the volume moving through each location will serve a wide market area. Figure 1 provides a summary of KMA assignments in the US.

Figure 1: Key Market Areas (KMAs) Map



(“Key Market Areas (KMAs) | SONAR Knowledge Center,” 2024)

1.2 Motivation

The optimal approach to investing in SCRE is complex but is highly dependent on the flows of goods from raw materials sources to finished product assembly and distribution. Transportation flows, as an extension of the global supply chain, have been more impacted by natural disasters, global events, and government regulations in recent years (TankTransport, 2024), which is important to our sponsor company as they look to invest in the best SCRE on the market. The sponsor wants to identify investment opportunities in advance based on new industry trends, rather than follow shifts in transportation flows after they happen.

In addition, the sponsor wants to evaluate any biases, conscious or unconscious, they may have regarding their preferred investment markets. Reasons for their bias could include familiarity and having previously invested in an area, reading news publications indicating supply chain activity in a certain area, or influence from industry peers that endorse investing in a specific location. The goal of this project is to evaluate markets for SCRE investment based on data and measurable factors rather than personal biases or anecdotal insights.

1.3 Problem Statement and Key Questions

The sponsor is looking to use data-driven methodologies to identify underserved markets in the US SCRE sector that will help guide their investment decisions for new properties. To identify investment opportunities, the sponsor needs to map the current state of US transportation flows, identify major regional trends in supply chain activity, and understand the gaps between the sponsor's current portfolio and future demand for SCRE.

Defining an underserved market is an important step for this project. To fully understand the concept, it's essential to break down the term "underserved market" into its two components – "underserved" and "market." The term underserved implies that an entity does not have access to adequate resources. Similarly, market can have different meanings, but for this project refers to geographic areas. The scope of this project is limited to US locations, so we aggregated the entire US into sub-markets by locational criteria by KMA.

The sponsor hypothesizes that a key method to identifying underserved markets is by tracking freight transportation flows in the US. This project sources freight flows data that is used to analyze the movement of goods in the US, including shipments' origins, destinations, and any stops in between.

Transportation flows geographies change as factors like demographics, industry practices, and consumer habits shift over time. It is crucial that the project captures trends in transportation flows over multiple time periods as it can provide insight into whether a market may be underserved.

Accordingly, this project aimed to answer research questions including:

- Where is freight moving to, from, and through in the US?
- How do we define an underserved market for supply chain real estate investments?

1.4 Goals and Expected Outcomes

During the project, we collected recent freight transportation data and identified major geographic markets in the US to develop an understanding of the current state of transportation

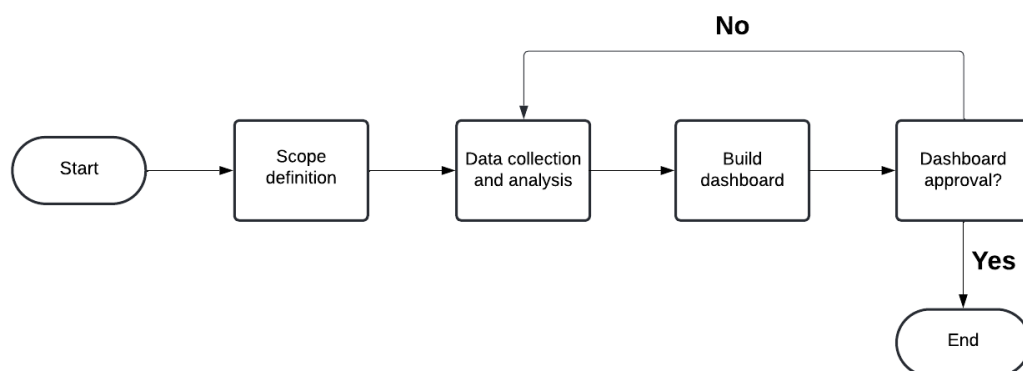
flows. To understand the shifts in freight transportation over time, we compared current data with historical data and identify key changes. Significant increases in transportation flows could imply the potential of new, major logistics hubs in the near or long-term future. Lastly, we overlaid the sponsor's current portfolio with current and historical data to identify underserved markets where there is an amount of supply chain activity greater than the amount of SCRE in the same geographic area.

The expected outcome of this project was a dashboard that visualizes the current state and trends in the SCRE sector and the sponsor company's portfolio. We included multiple data sources in our analysis, including transportation flows data and SCRE property details. The data inputs can be dynamic and regularly updated, so the dashboard can evolve as new iterations of the data are released.

1.5 Plan of Work

We completed multiple versions of the dashboard before it was finalized. We regularly solicited feedback from our sponsor company to ensure the dashboard is in a clear visual format that shows all requested data points like transportation flows and property locations. We remain open to including new data sources and corresponding analysis in the dashboard. Figure 2 below shows the key steps that were completed for this capstone project.

Figure 2. Plan of Work



2. STATE OF THE PRACTICE

This capstone project required a deep understanding of the current global logistics landscape, SCRE, and transportation flows. This chapter highlights relevant research in these topics and explains how they are related to the overall goal of the project. Industry-leading research on SCRE and transportation flows in the US are important to consider as they provide analytical and strategic insights that inform the project's methodology and objectives.

2.1 Logistics Trends Impacting Supply Chain Real Estate

The supply chain and logistics sectors continue to evolve as companies respond to various disruptions around the world like the pandemic, tariffs, and changing consumer habits. Trends like nearshoring and e-commerce growth will continue to impact companies' global operations and are relevant to our capstone sponsor's SCRE portfolio.

In recent years, US-based companies have nearshored some of their supply chain operations from outside North America to the US and nearby Canada or Mexico. Industry experts estimate that Mexican imports into the US could increase by 40% over the next five years, with a large proportion coming through Texas inland ports like Laredo. Some third-party logistics providers have increased their operational presence in the area (Wolf, 2023). Certain US ports, like those in Texas or California that border Mexico, are best equipped to receive imports if this trend continues.

As e-commerce continues to grow, so will the demand for final-mile delivery facilities and services. The global e-commerce market is projected to grow at a nearly 15% rate compounded annually. This comes after worldwide e-commerce sales have tripled in the past seven years (Bledsoe, 2024). Public and private entities will continue to invest in final-mile delivery facilities to meet this demand for e-commerce services.

2.2 Supply Chain Real Estate

Supply Chain Real Estate (SCRE) is a component of industrial real estate and in our case refers to properties used for warehousing, storage, and distribution of goods throughout the US. The location of such properties is often determined by transportation flows in the US. SCRE should be in convenient locations for truck movement. This means it should be near highways and along major routes that trucks will have to take from their origin to their destination. Proximity to

highways is important so that logistics providers can minimize both the time and costs of traveling to a facility.

SCRE has grown increasingly important over the years as the impact of global events has taken a toll on the supply chain. More companies are considering near or onshoring production (Alvim & Averbuch, 2023) and the increase in e-commerce means that the need for properly located SCRE is not going away any time soon. In fact, transportation flows in the US are projected to nearly double between 2023 and 2050 (Elgin, 2024). Doubling the volume moved via truck in the US means the SCRE industry must expand to meet the capacity demanded. SCRE should be invested in key areas across the US that face the greatest surges in transportation flow volume.

The sponsor company's business model is to purchase SCRE and lease out the properties to tenants who pay rent. Property characteristics like number of doors, building age, and office space availability tend to be linked to the price to lease a building. Additionally, macroeconomic factors like GDP growth have a positive correlation to rent prices due to increase of demand (Benjamin et al., 2003). Multiple factors contribute to leasing price and demand of SCRE which directly impact the sponsor's revenue.

A previous capstone with the same sponsor company was conducted similarly to ours to identify underserved geographic markets that warrant additional SCRE investment. The project used ordinary least squares regression, regularized regression, and XGBoost models that incorporated transportation flows and metrics that are indicative of the trucking and US economy conditions (Hernandez & Huisman, 2024).

2.3 Transportation Flows Impacting SCRE

It is extremely important for our sponsor to understand the factors affecting industrial properties' valuations to make proper investments. Research indicates that location is the most significant determinant, accounting for 24%, while physical characteristics, economic factors, and transportation infrastructure each contribute over 10% (Razali et al., 2020). While general market location is extremely important, the price can vary with land near urban areas higher than land near rural areas (Razali et al., 2020). Determining whether to purchase more urbanized or rural land is the task of the sponsor who has the expertise and understanding of what would be most attractive to a given customer. However, we are making pivotal contributions by highlighting key markets for the sponsor to investigate further.

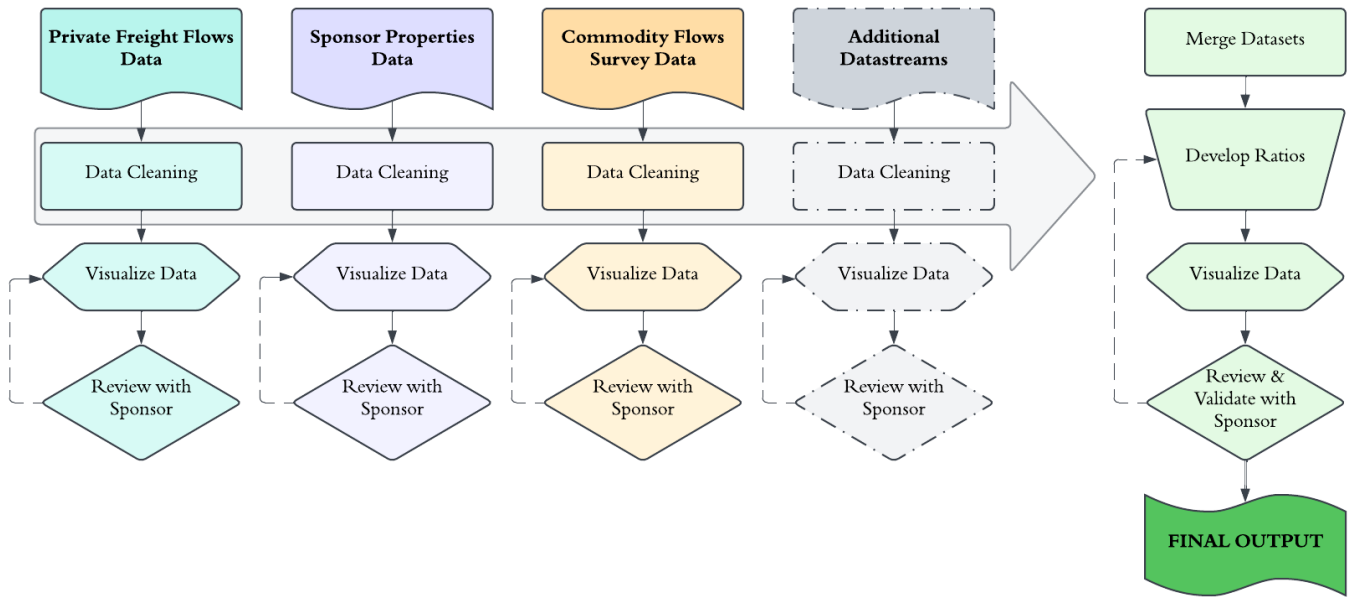
Transportation flows help determine the ideal locations of SCORE. However, transportation flows are extremely difficult to predict. Many studies, as well as alternative capstones, have been conducted to predict US trucking flows as well as boom and bust cycles to predict shifts in major logistics hubs (Acocella et al., 2020). These studies use a variety of analytical methods, like linear regressions and algorithmic modeling. They use external factors such as population shifts, fuel prices, and consumer habits as independent variables to predict trucking flows.

This project does not use advanced analytics like regression, machine learning, or deep learning but rather provides a highly functional and user-driven dashboard that enables the sponsor to manipulate and extract insights from the data. The goal is for the sponsor to receive a tool that displays the current state of transportation flows overlaid with properties, both owned and investigated, and visual cues that inform them of investment opportunities at the KMA-level. With a simplified methodology, we ensure the sponsor can interpret the dashboard and update its data inputs as more information becomes available. The dashboard is also highly customizable and can be tailored for specific insights as the sponsor continues to use it.

3. METHODOLOGY

This section summarizes our methodology for analyzing the sponsor's portfolio, integrating US transportation flows data, and defining underserved geographic markets. Figure 3 details the general process of the data cleanse, data visualization, and ratio development steps followed. This methodology is an iterative process and can be expanded to include additional datasets beyond what we incorporated, should the sponsor decide to pursue more granular and/or up to date inputs.

Figure 3: Methodology Process Flows



Built using LucidCharts. Colors serve as a callout for different data sources. Dashed arrows imply optional loops. Dashed boxes indicate optional additional steps.

3.1 Collect Data on Transportation Flows

The project began with an assessment of overall transportation flows in the US and an analysis of trends affecting companies' supply chain operations. There were both public and private data options available for this project. They vary in cost, scale, and years available. The priorities were to find low/no-cost data, that represents the larger market and was compiled from recent years.

Initially, the US Commodity Flows Survey (CFS) from U.S. Census Bureau was considered. This survey's results are free and accessible to all. It includes information on transportation of goods throughout the US and categorizes them by material type, mode of transport, weight, origin, and destination (United States Census Bureau, 2024). However, the CFS is only released every five years, often with further delays. The most recent CFS was released in September 2020 and only includes data through 2017. The latest version is planned for release later this year. CFS data was included in our analysis, but it is meant to be supplementary due to older dates and lower relevance.

Many private data sources are available; however, they are either protected as confidential information or sold at a high price. They do provide important details of transportation flows like the CFS but tend to represent small portions of the broader trucking industry. What sets private data sources apart though, is they are often updated more frequently and/or cover more recent time periods (only a few weeks/months delay).

In the case of this project, a digital freight broker (DFB) provided data on their historic bookings. The DFB serves multiple industries. They operate in all continental US states with freight services like full truckload, LTL, and drayage. DFB's vast transportation and logistics service offerings relate to the sponsor company's properties, as shipments facilitated by the DFB would feasibly move through a facility the sponsor company owns. The transportation flows data used covers over 200,000 bookings, which is a small fraction of the total transportation flows in the US but provides a proxy to the overall movement of goods in the US.

The DFB data covers the years 2019 through part of 2023 and includes metrics like total weight moved and count of contracts awarded to origin and destination KMA lanes. These years incorporate periods before, during, and after the COVID-19 pandemic when supply chains across the globe were disrupted.

The use of this private DFB data is representative of the broader logistics industry and covers multiple recent years. Should the sponsor want to expand on the data sources used for this project, we recommend they add up-to-date transportation flows data and property information into the dashboard so the insights from the tool do not become obsolete. This capstone acts as the catalyst for the sponsor to view SCRE in the US differently and is part of the foundation for their future strategic property investments.

3.2 Analyze Transportation Flows Data

After receiving the data from the CFS and the DFB, we assessed their validity, which included confirming that each record had a proper origin and destination city, correctly entered weights, and properly assigned ZIP codes. Accurate analysis would have been difficult to conduct without consistency in the data sets and lead to unreliable insights. Validation was completed using multiple methods ensuring city, state, latitude, longitude, and/or other descriptions of each property matched.

Once the data was validated, ZIP codes were matched to KMAs. KMAs allow for simpler aggregation across all ZIP codes in the US so that the dashboard visualizes 135 KMAs rather than tens of thousands of individual ZIP code locations. KMAs aggregate three-digit ZIP codes. The output for each KMA includes an average latitude and longitude. For example, KMA “AL_BIR” covers the greater Birmingham, Alabama, area and includes three-digit ZIP codes 350, 362, 352, 351, and 354. The average latitude and longitude are 33.608 and -86.863 respectively.

All data cleaning and analysis were done using Python, a general-purpose programming language. In theory, data cleansing could have been done in Microsoft Excel, but the time commitment and validation process would be much larger and more prone to error due to the sheer size of the data sets.

We were able to summarize flows at the KMA-level by inbound, outbound, and total due to origin and destination pairings. In the data, anything leaving an origin KMA is an outbound shipment and anything entering a destination KMA is an inbound shipment. The sum of both created the total flows which are a key data point in the project. With each transaction assigned to a KMA, the resulting data set consisted of 135 rows, one for each KMA, with sum of inbound, outbound, and total flows, as well as average latitude and longitude.

Based on the DFB’s transportation flows data in 2022, the Atlanta, GA, KMA has the highest total volume (5162), followed by Columbus, OH, (4423) and Dallas, TX (4379). GA_ATL also has the highest inbound volume (2798), while OH_COL (2920) has the highest outbound volume. The median total volume is 560, but average is 965, indicating that the data is right skewed with the larger values pulling the mean higher.

A brief overview of the high volume KMAs in the DFB data is shown in Tables 1-5. The variability in volume shows the difficulty of predicting trucking flows in the US. There are clear imbalances across the regions which make it even more of a challenge for our sponsor to determine where to invest next. Additionally, there are significant shifts in flow imbalance as certain KMAs see much more inbound volume than outbound volume, and vice versa.

Table 1: DFB Top KMAs Based on Total Volume

Highest Total Volume
Atlanta, GA
Columbus, OH
Dallas, TX
Ontario, CA
Charlotte, NC
Allentown, PA
Harrisburg, PA
Chicago, IL
Elizabeth, NJ
Grand Rapids, MI

Table 2: DFB Top KMAs Based on Inbound Volume

Highest Inbound Volume
Atlanta, GA
Dallas, TX
Allentown, PA
Ontario, CA
Charlotte, NC

Table 3: DFB Top KMAs Based on Outbound Volume

Highest Outbound Volume
Columbus, OH
Atlanta, GA
Charlotte, NC
Dallas, TX
Elizabeth, NJ

In Tables 1-3, the top five KMAs in Table 1 are not exactly aligned to the top five in Table 2 or Table 3. This shows a clear imbalance in inbound and outbound volumes even among the top KMAs for truck flows.

Table 4: DFB Top KMAs Based on Total Volume

Highest % Inbound*
Albuquerque, NM
Spokane, WA
Tifton, GA
Montgomery, AL
Austin, TX

**Minimum volume of 100.*

Table 5: DFB Top KMAs Based on Total Volume

Highest % Outbound*
Buffalo, NY
Taylorville, IL
Rochester, NY
St. Louis, MO
Medford, OR

**Minimum volume of 100.*

Tables 4 and 5 show the KMAs with the biggest imbalance in freight flows. The top five for either table have greater than 70% of inbound or outbound flow respectively.

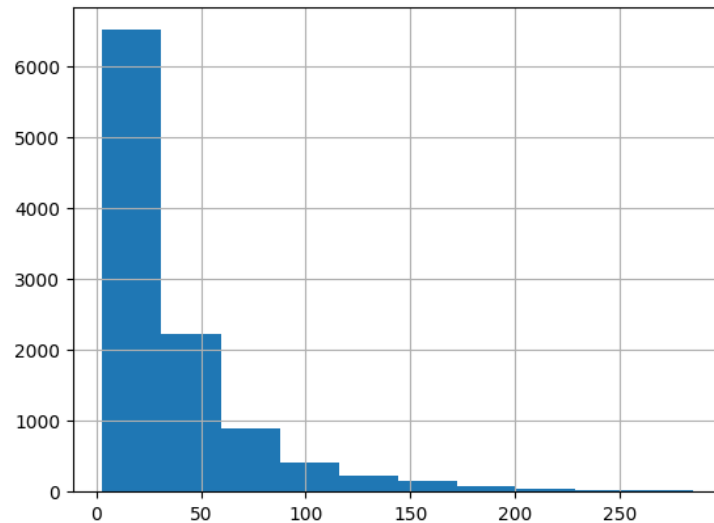
3.3 Assess Sponsor Company's Portfolio

The sponsor has various properties across the globe as well as extensive details on properties of interest. They do not own properties of interest but track them to better understand what is active in each market. Overall, there are more than 30,000 properties included in the data, but the sponsor does not own all of them. Our capstone does not differentiate between owned and tracked properties, but this would be another layer to add to the data in future iterations of the dashboard.

For each record, there is some combination of location area, acreage, loading positions, parking spaces, offices, and other key features. We aggregate all values at the KMA-level for US properties but keep track of the number of properties with Record Count. The sponsor can use this information as well to better understand if some markets are catered towards fewer, massive properties or more, small properties. Additionally, we removed any outliers in the data that had total loading positions less than 2 and greater than 300. This range was agreed upon with the

sponsor as excluding any error-prone entries (negative loading positions), tiny properties, or massive properties that are not common investments. Distribution of properties and loading positions can be seen in Figure 4.

Figure 4: Distribution of Property Loading Position Greater Than 2 and Less Than 300



In cleaning the property data, the first task was isolating properties located in the US as international properties are not in scope. Like the transportation flows, having accurate latitudes and longitudes of each property was essential, as coordinates are used to map properties in the visualizations. For many properties, latitude and longitude were missing, switched, or inaccurate. In these cases, a combination of city, state, address, and ZIP code information was used to segment the data. After the US properties were isolated, each was assigned a KMA based on three-digit ZIP code and information was aggregated.

The data provided did not explicitly specify which properties were truck terminals, secure storage and parking, and final-mile delivery facilities. Each property was assigned an HFT Property type based on its description and discussion with sponsor as seen in Table 6. For example, properties described as “LTL Cross Dock” were assigned to truck terminal and “Drayage Yard” to secure storage and parking. This helped narrow the property types into a few distinct categories rather than numerous descriptions.

Table 6: HFT Property Mapping

HFT Property	Original Property Types
Secure Storage & Parking	• Shop • Truckload Yard • Drayage Yard
Final Mile Delivery	• Final-Mile Warehouse • Warehouse • Parcel Building • Cold Storage • Food & Beverage
Truck Terminal	• LTL Cross Dock • Transload Facility
Other	• Land

Understanding which features and metrics in the property data most strongly contribute to the value of each SCRE site is crucial. We decided to focus on loading positions, meaning truck bays in a facility, as a primary property feature, because the number of truck bays in a facility indicates the demand of trucks each facility can support. Once the data was segmented and aggregated, there were approximately 10,000 properties remaining spanning all KMAs with an average amount of loading positions per property of 58.

3.4 Build Visualizations

Power BI was the tool used for visualizations in this capstone. The final output maps both transportation flows and the sponsor company's portfolio of properties. This tool is particularly useful because of its geographic and spatial capabilities. It has built-in features like ArcGIS and Azure Maps that accurately map data points based on coordinates. Additionally, the maps are highly interactive and can be synced for quick comparisons and deep dives. Lastly, Power BI easily integrates multiple data sources like Microsoft Excel, SQL, and Text/CSV. The sponsor company is in the process of integrating Microsoft Fabric into its resources which will allow the company to link and automatically update their property data, CFS data files, and DFB/other private data files seamlessly in the future.

The initial work with visualizations involved multiple trial dashboards and we regularly solicited feedback from the sponsor based on ease of use, readability, and insights. This not only drove the discussion on the sponsor's bias in what market prioritization, but it also helped us develop the most impactful visuals. Versions of dashboards included total freight volume, inbound volume, and outbound volume by KMA. The sponsor responded positively to the use of color scales and size differences to distinguish flows and markets. Examples of the initial simple visuals can be found in Figures 4, 5, and 6. Figure 4 shows a map of total shipment volume, both inbound and outbound, by KMA in the US. Figures 5 and 6 show inbound and outbound shipment volume by KMA. The size of each bubble represents the aggregate shipment count by a KMA and is mapped to its geographic location. All three of these figures use data only from the DFB.

Figure 5: Total Volume by KMA

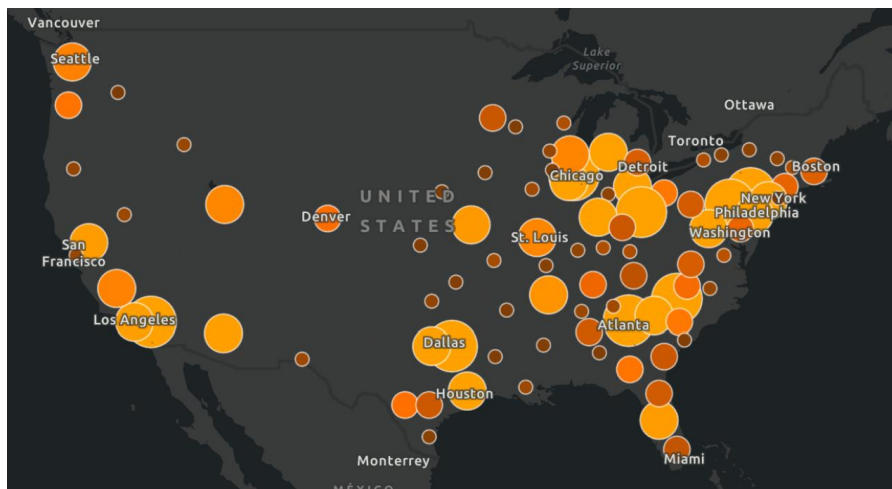


Figure 6: Inbound Volume by KMA

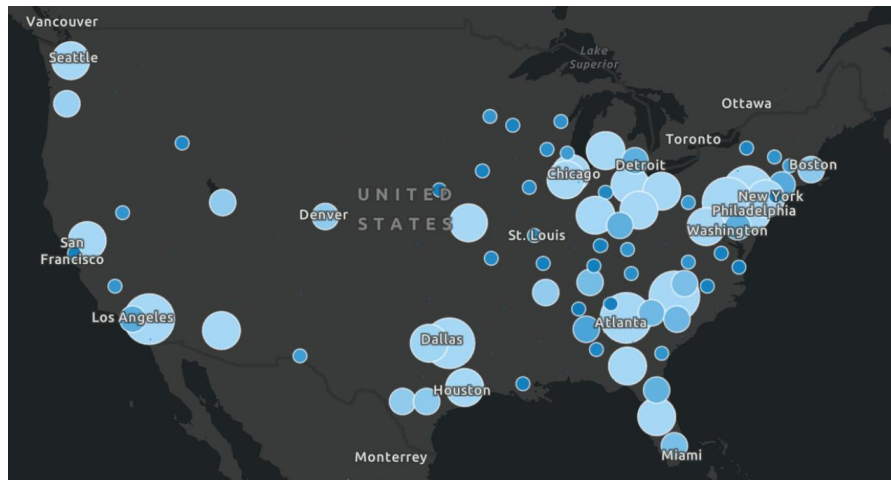
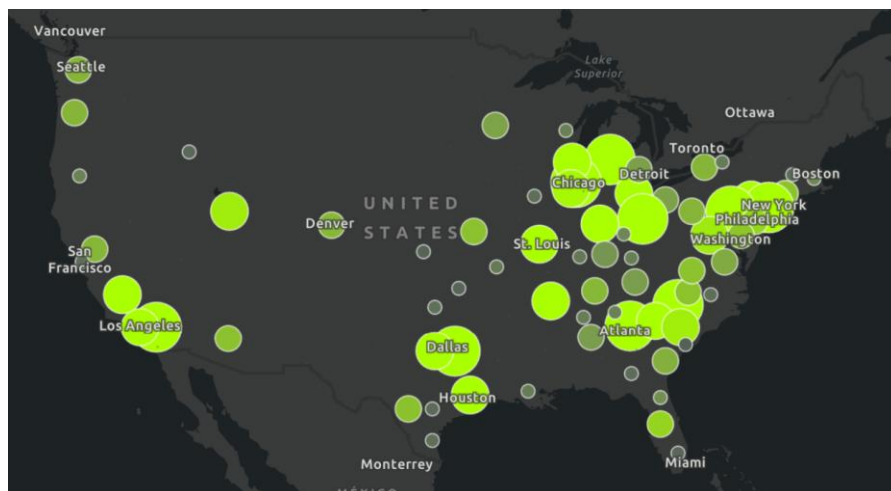


Figure 7: Outbound Volume by KMA



Using a similar mapping approach, the team visualized the sponsor's properties by total acreage, number of loading positions, and total area by KMA. Loading positions were selected as the primary data point, but all other property specifications were kept as additional information the sponsor can consider in their analysis.

The most advanced dashboards created for this project include comparison of transportation flows and property data. The merger of transportation flows and sponsor property data into an easily interpretable and highly interactive dashboard was the key deliverable of this project and is covered in more detail in the results section of the report.

3.5 Merge Data Sources and Develop Ratios

After testing separate visualizations with only DFB trucking flows and sponsor property data, the next step was creating a merged file combining all data sets that could be used in the dashboard visualization.

With the data fully merged into one file at the KMA-level, we defined ratios of Loading Positions / Total Flows from DFB data and Loading Positions / Total Weight from CFS data. While both datasets measure trucking flow movement throughout the US, they do so on different scales as the DFB data uses the amount of shipping transactions while CFS data counts total weight. The CFS ratios were normalized to the scale of the DFB ratios to ensure they could be easily compared. A min-max normalization was applied from the DFB data to the CFS data in a standardized 0–1 range. Then the normalized values in the CFS data was rescaled to the empirical range of the DFB ratios. This created a set of adjusted CFS values that maintained their relative position while being comparable to the DFB values.

Table 7 summarizes the primary data used in the dashboard and how it is aggregated. Inputs from both the transportation flows and sponsor data are specified. Additionally, calculations that were made using the transportation flows and sponsor data are identified.

Table 7: Primary Data Used

Column Name	Data Source	Purpose	Aggregation Level
KMAs	Truck Flows Data	Key Market Areas that data is aggregated to for ease of data manipulation and visualization.	KMA-level.
HFT Property	Sponsor Data	Simplified version of sponsor property types. Adds options for segmenting the data based on specific property type(s).	KMA-level.
Latitude	Sponsor Data	For geographic location on visuals.	Averaged at KMA-level.
Longitude	Sponsor Data	For geographic location on visuals.	Averaged at KMA-level.
Loading Positions	Sponsor Data	Number of “doors” or truck bays at a given property. It is an indicator of capacity capability at a site.	Sum at KMA-level.
Year	Truck Flows Data	Allows for year-over-year comparisons	KMA-level.
Flow Imbalance	Calculated (for DFB and CFS)	Inbound Flow (or Weight) minus Outbound Flow (or Weight). Imply any imbalance in truck movement through each KMA.	Sum at KMA-level.
Record Count	Calculated	Defines the number of properties in each market.	Count at KMA-level.
Flow Ratio	Calculated (for DFB and CFS)	Calculated as Loading Positions / Total Flow (or Weight). Primary Ratio used to define underserved markets.	KMA-level.
Ratio Rank	Calculated (for DFB and CFS)	Rank KMAs in each year based on Flow Ratio. 1 being the highest total flow, 132 being lowest.	KMA-level.
Quantiles	Calculated (for DFB and CFS)	Used to sort and categorize data as well as define underserved markets.	KMA-level.

Other data regarding sponsor properties was kept in the merged file for their use in analysis of underserved markets. When the sponsor interacts with the dashboard, they will have access to additional information for each KMA by hovering over a given bubble or selecting in a table. A sanitized version of the final dashboard can be found in Results as Figures 8-10.

3.6 Quantify Underserved Markets with Merged Data

Integrating the transportation flows and property data is key to defining an underserved market. As stated earlier, underserved markets are geographic areas where there is proportionally more supply chain activity than the SCORE available in the same area. However, there are no specific guidelines as to what difference in proportion would be considered underserved, nor the best way to quantify supply chain activity and SCORE for a proper comparison.

The project team and sponsor hypothesized that the ratio of freight volume to loading positions would be a leading indicator of an underserved market. Different ratios were also calculated

including total area to freight balance (inbound minus outbound), total acreage to freight volume, and loading positions to freight volume, all at the KMA-level. We provided all ratios to the sponsor but focused on total flow and total weight ratios for our analysis.

The final dataset and dashboard included ratios over years, breakdowns by HFT Property, and high-level overviews of DFB and CFS flows. The sponsor can compare the 2017 CFS data and ratios with 2020, 2021, or 2022 DFB data and ratios. It can easily filter by year as well as property type if they want to focus on one or more specific property type for investment.

The dashboard also includes actively calculated quantiles of the ratios. The quantile values used are 0.1, 0.2, 0.5, 0.8, and 0.9. The respective ratios can be seen in the Results section in Table 12 and compared between the DFB data and normalized CFS data.

4. RESULTS

In this section, we first present the outcome, the interactive dashboard that our sponsor requested. We will then summarize the findings regarding the definition of an underserved market as well as multiple ways the sponsor can use this tool in determining the next market to invest in. We then discuss the overall implications of this tool and then summarize the limitations and future areas of improvement.

4.1 The Interactive Dashboard

In Figures 8 and 9 below, there are snapshots of the final dashboard delivered to our sponsor. In the top half, it includes a table and a visual combining the CFS data with the sponsor property data. In the bottom half, it includes the same for the DFB data merged with the sponsor property data.

Each map has varying bubble sizes for each KMA. A larger bubble signifies a higher trucking flow by count (DFB data) or by weight (CFS data). You will see in the maps in Figures 8-10 are monochromatic to keep the data provided confidential. In the actual output, there is a color scale associated with the ratios for every KMA. Additionally, the tables will normally include the KMA name (e.g., AL_BIR), but we removed those as well to keep the information confidential. On the right-hand side of the dashboard, the sponsor can filter on a year of DFB data as well as

the types of HFT property data they want to compare against. The dashboard automatically updates ratios, rankings, and quantiles based on the filters.

Figure 8: Final Dashboard with Highest Ranked KMAs



Figure 9: Final Dashboard Highlighting Lowest Ranked KMAs (Likely Underserved)

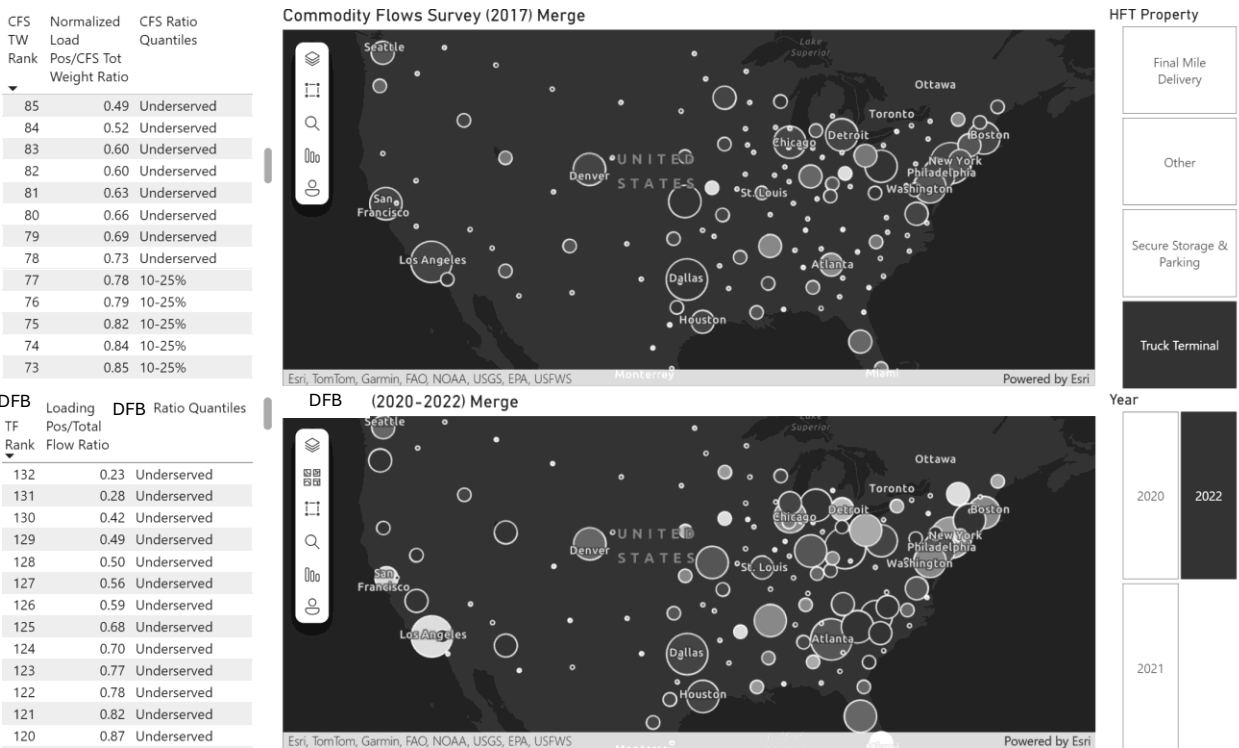


Figure 10 below provides an example of the ease of use and interaction with the dashboard. The user selected the 79th ranked KMA in the CFS table in the top left, which automatically filtered to the same KMA in the DFB table as well as focused each map on the specific KMA region. If the user were to hover over one of the bubbles, they would see a Power BI “ToolTip” which is an information box that provides all the additional details about the properties and KMAs we have. Each tooltip includes total loading positions, total flow/total weight, flow/weight imbalance, acreage, and more beyond the KMA and its assigned ratio.

Figure 10: Final Dashboard Highlighting Interactive Map



Underserved markets were calculated using both the transportation flows and sponsor data. The ratio of total loading positions to total transportation flows by KMA is used as the primary result to determine underserved markets for this project. A lower ratio indicates that a market may be underserved.

4.2 Ratio Results

Ratios to measure whether a market is underserved range from 0.23 to 9.78. Table 8 and Table 9 shows the KMAs with the lowest ratios for CFS and DFB data, while Table 10 and Table 11 show the KMAs with the highest ratios. The name of each KMA has been removed and only rank remains in each of the tables below. Each ratio can be interpreted as the number of doors relative to a given unit of truck volume (weight for CFS data or contract count for DFB data).

Table 8: CFS Lowest 5 Ratios (Possibly Underserved Markets)

Rank	Normalized Loading Position / Total Weight Ratio
85	0.49
84	0.52
83	0.60
82	0.60
81	0.63

Table 9: DFB Lowest 5 Ratios (Possibly Underserved Markets)

Rank	Loading Position / Transportation Flows Ratio
132	0.23
131	0.28
130	0.42
129	0.49
128	0.50

Note that the datasets have different numbers of KMAs, 86 KMAs for CFS and 132 KMAs for DFB. While the federal CFS is a data source with many more data points than the private DFB data, it does not encompass as many KMAs as the private data source. We believe this is still extremely helpful for the sponsor's use and can simultaneously highlight different data sources that are consistently important KMAs with high or low ratios as well as help the team investigate any significant differences between data sets.

Table 10: CFS Highest 5 Ratios

Rank	Normalized Loading Position / Total Weight Ratio
1	9.78
2	6.12
3	5.52
4	4.34
5	3.57

Table 11: DFB Highest 5 Ratios

Rank	Loading Position / Transportation Flows Ratio
1	9.78
2	6.85
3	6.59
4	6.50
5	6.48

4.3 Interpreting Ratios

Rationally, a ratio of 1 or less may suggest an underserved market, as it indicates there is only one or fewer truck bays available per truck. However, considering that the ratio is subjective in meaning, we wanted to add structure to how we defined underserved as well as how we prioritized markets that we believe are underserved.

We determined with the sponsor that any ratio at or below the 10th quantile (0.1), would be considered underserved and a priority for additional research for an investment. Additionally, any ratios less than 50th quantile may need to be reviewed or investigated to see if it is also underserved. The additional research and review incorporates the SCORE expertise that the sponsor company has from years of experience and historical investment that we as the project team do not. By using this merged data in combination with a broader understanding of the market, the sponsor removes initial bias towards investing in certain markets, assesses markets deemed underserved that they would not normally consider, and then prioritizes markets in an investment strategy.

Table 12 compares DFB and CFS ratio quantiles. In the red, dashed box we identify the quantile that is highly likely to be underserved. Any KMA with a ratio at or below that value should be investigated immediately for investment opportunities. In the orange, dotted box, we identify any ratio above the 10th but below the 50th percentile. Any KMA with this ratio should be reviewed at a secondary stage to access investment opportunities.

Table 12: Comparison of Ratios at Each Quantile

Quantile	DFB Loading Pos/Total Flow Ratio	CFS Normalized Load Pos / Total Weight Ratio
0.1	1.329	0.770
0.2	1.599	0.899
0.5	2.812	1.092
0.8	4.250	1.942
0.9	4.964	2.851

All this information is readily accessible in the dashboard visualization and the sponsor can complete deep dives on each KMA in a matter of seconds to create a list of markets to prioritize.

4.4 Contributions to Sponsor

Our analysis can be used to support SCRE investment decisions for the sponsor. In prior investment decisions, the sponsor focused solely on historically important markets, without considering some of the additional 100+ KMAs across the US. The dashboard combines multiple data sources of trucking flows to show exactly where freight is moving to, from, and through in the US. It then combines that data with sponsor property information and merges the two using a ratio of Loading Positions to Truck Flows.

In combination with additional real estate investment insights like historic market understanding, property values, local laws and regulations, and other due diligence, the sponsor now has a tool to guide them in future investments. The sponsor can manipulate and draw insights from the tool as they see fit. This includes double-clicking on a KMA to see its transportation flows and

industrial real estate data or filtering for certain property types like truck terminals or final mile delivery facilities. Additionally, the sponsor can use the two different data sources, Federal through CFS and Private through DFB, to compare flows by years as well as the differences between the two data sets.

This tool and methodology is the first of its kind for our sponsor as well as a novel way of thinking about SCORE. We have already received feedback from the team on multiple KMAs they were surprised by and wanted to further deep dive investment opportunities.

Additionally, how the sponsor thinks about their property data and market analysis has changed after completion of this project. The use of straightforward ratios to merge property data with trucking data has not been done before by the team. It opens a door to several possibilities for future data collection. Whether they want to update with the latest CFS data that will come out later this year or incorporate additional sources, all can be easily integrated into the merged data source and used to add more layers of context to why trucking flows move to, from, and through given markets as well as how exactly the sponsor should prioritize those markets for investment.

4.5 Limitations and Areas of Improvement

Data scale, quality, and recency are all key to the development and improvement of this capstone. Underserved market analysis can be expanded by including additional and more up-to-date data.

Our capstone is limited by free data sources with the latest date for trucking flows being 2022. Many other private data sources, such as DAT Freight and Analytics could be used to improve the recency of the data as well as adjust the scale/granularity if the sponsor wants more customization. DAT is a software company that provides information of trucking movement on a much more regular basis. This data comes at a cost, however, and we suggest the sponsor continue using our visualization and data set until they identify exactly what specifics would help them in the future.

On the sponsor data side, metrics like property utilization data would be useful to determine if a KMA is underserved or overserved. The sponsor could see year over year how many of their properties are actively in use at a particular time and how they change depending on shocks or shifts in trucking flows. We do not currently have information on the number of actively in-use

properties in the dataset. A lack of filter on property utilization may cause the ratio of a given KMA to be higher, because realistically a portion of its aggregated properties, and therefore, loading positions, are not currently leased and in use. Additionally, the analysis does not segment sponsor-owned vs. market-monitored properties. Differentiating based on these criteria would allow for analysis on the sponsor's actual portfolio gaps. If they identify that a market is not only underserved, but there are multiple properties available that they do not own, that provides a clear path forward in investments.

Lastly, due to Power BI functionality, we cannot do any auto-updates and linking between the dashboard and the cleaned data output from the Python code. For every new iteration of the merged data, the file must be downloaded, added to the Power BI dashboard, and replace the old file version with the visuals. This is a task that can be automated within the sponsor's systems. They are undergoing an integration of Microsoft Fabric, an enterprise-level, cloud-based analytics platform. By incorporating the merged data and visual into the Fabric overhaul, the sponsor can then complete faster updates which allow for more iterations of the visuals should they decide to include additional or different data sources.

5. CONCLUSION

The goals of this project were to answer two questions.

- Where is freight moving to, from, and through in the US?
- How do we define an underserved market for supply chain real estate investments?

This project began with an analysis of transportation flows in the US and the SCRE markets that the sponsor operates in. Then, all data was aggregated at the Key Market Area (KMA) level to condense the analysis from thousands of ZIP codes to 135 standardized geographic markets. The data was then incorporated into an interactive and visual dashboard that shows both truck flows through the US and the sponsor company's property analysis. Lastly, ratios were calculated that compared transportation flows with loading positions (doors) of all SCRE properties within a KMA. These ratios provide data-driven and bias-free insights into determining underserved markets and give the sponsor a method to prioritize their next investments.

5.1 Insights and Recommendations

We recommend the sponsor use the dashboard developed in this capstone as a tool to define and prioritize target markets to invest in. It could be referenced on a quarterly or annual basis to define the next period's strategy.

Including additional data points like property occupancy status and regularly updating information in the properties database would be beneficial to the sponsor company. This would allow for additional insights into whether strategic investments are translating into active utilization, which is the main revenue driver for the sponsor.

5.2 Future Applications

The sponsor now has a tool to visualize trucking flows in the US from two different sources as well as identify underserved markets to drive their investment decisions. The dashboard will be a significant contribution to the sponsor as it provides a data-driven framework that identifies mismatches between transportation flows and SCRE infrastructure within geographic areas. We suggest that the sponsor uses this dashboard and all the project's analysis as part of their due diligence for future investment decisions.

Additionally, this methodology could transfer to other countries the sponsor invests in. SCRE is required all over the globe, but trucking flows and accessibility of data vary by region and country. Our methodology is simple, scalable, customizable, and at no cost to the sponsor. After using it to develop the US investment strategy, the sponsor may identify data sources in other regions that could also be compared with properties and loading positions in those areas. While there are not as clear cut KMAs in all countries as there are in the US, this capstone opens a world of investment opportunities for the sponsor.

REFERENCES

- Acocella, A. (2020). *Wayfair Inc.: The Art and Science of Transportation Procurement*.
- Acocella, A., Caplice, C., & Sheffi, Y. (2020). Elephants or goldfish?: An empirical analysis of carrier reciprocity in dynamic freight markets. *Transportation Research Part E: Logistics and Transportation Review*, 142, 102073.
- Alvim, L. & Averbuch, M. (2023, June 28). *US Nearshoring Wave Grows as Mexico Exports Jump Close to Record*. Bloomberg. <https://www.bloomberg.com/news/newsletters/2023-06-28/supply-chain-latest-us-nearshoring-proof-grows-as-mexico-exports-jump>.
- Benjamin, J. D., Zietz, E. N., & Sirmans, G. S. (2003). The environment and performance of industrial real estate. *Journal of Real Estate Literature*, 11(3), 279–321. <https://doi.org/10.1080/10835547.2003.12090130>
- Bledsoe, J. (2024). *E-commerce sales size & forecast*. U.S. Department of Commerce, International Trade Administration. <https://www.trade.gov/ecommerce-sales-size-forecast>.
- United States Census Bureau (2024). *Commodity Flow Survey (CFS)*. Census.gov. <https://www.census.gov/data/developers/data-sets/cfs.html>.
- Cauble, T. (2023, September 1). *The Ultimate Guide to Industrial Real Estate*. The Cauble Group. <https://www.tylercauble.com/blog/industrial-real-estate>.
- Elgin, S. (2024, August 20). *Trucking Industry Trends, Statistics, & Forecast – 2024 Edition*. Truck Info. <https://www.truckinfo.net/research/trucking-statistics>.
- Hernandez, S. & Huisman, S. (2024, May 10). *Data-Driven Strategies for Identifying Underserved Markets and Optimizing Real Estate Investment in the U.S. Transportation Sector*. Massachusetts Institute of Technology. <https://ctl.mit.edu/pub/thesis/data-driven-strategies-identifying-underserved-markets-and-optimizing-real-estate>
- Jacobs, K. J. (2021). Build-to-Suit vs. Spec: Which Building is Right For a Specific Company?. *NAIOP | Commercial Real Estate Development Association*. <https://www.naiop.org/Research-and-Publications/Magazine/2021/Summer-2021/Development-Ownership/Build-to-Suit-vs-Spec-Which-Building-is-Right-For-a-Specific-Company>.
- Jenkins, A. (2023, September 7). *Inbound vs. outbound logistics: Key differences and best practices*. NetSuite. <https://www.netsuite.com/portal/resource/articles/inventory-management/inbound-outbound-logistics.shtml>.
- Maersk. (2024, July 5). *Drayage: A key part of supply chains*. Maersk. <https://www.maersk.com/logistics-explained/transportation-and-freight/2024/07/05/drayage>.
- Nahata, K. (2022, September 2). *Managing the complexities of last-mile delivery*. Forbes. <https://www.forbes.com/councils/forbesbusinesscouncil/2022/09/02/managing-the-complexities-of-last-mile-delivery>.

Pitz, P. (2017, January 18). *Free Zip Zone and KMA Wall Map Available - DAT Freight & Analytics - Blog*. DAT Freight & Analytics. www.dat.com/blog/free-zip-zone-and-kma-wall-map-available.

Prologis. (2019, September 17). *The Modern Supply Chain: A New Model for Defining Logistics Real Estate*. Prologis. www.prologis.com/insights/global-insights-research/modern-supply-chain-new-model-defining-logistics-real-estate.

Rodrigue, J.-P. (2024). 5.1 – Transportation Modes, Modal Competition and Modal Shift. In *The Geography of Transport Systems* (6th ed.). essay, Routledge.

Razali, N. S., Menan, N. I., Teng, S. V., Ting, N. L., Mahayuddin, N. H., & Rahman, M. S. A. (2020). Factors Affecting Industrial Property Value. *International Journal of Scientific & Technology Research*, 9(01), 212–217.

Sonar. (n.d.) *Key Market Areas (KMAs)* Sonar. <https://knowledge.gosonar.com/resources/sonar-data/trucking/key-market-areas-kmas>.

Tank Transport. (2024, November 26). *7 ways global events impact trucking*. Tank Transport. <https://tanktransport.com/2024/11/7-ways-global-events-impact-trucking>.

Wolf, C. (2023, September 13). *Nearshoring Boom Faces Growing Pains*. Transport Topics. <https://www.ttnews.com/articles/nearshoring-growing-pains>.