August 13, 2024

High-Speed Passenger Rail

Presented by Team 54

Kevin Houseman, Tori Sosnowski, Leslie Stovall, Emma Valind

Photo by Charles Forerunner on Unsplash

Table of Contents

Executive Summary

Problem Statement

High-speed trains have successfully driven economic growth, created jobs, and brought communities together around the world. These benefits have remained largely untapped in the United States, which remains reliant on auto and air travel. However, the U.S. has recently begun reinvesting in its railway networks and allocated funds for the development of high-speed rail.

In order for these new high-speed rail projects to maintain public support, it is crucial that they are seen as successful. **A key factor of this success is identifying priority corridors for high-speed rail development**. Prioritizing areas with the right mix of population density, economic activity, and existing travel demand helps ensure strong ridership numbers and reduces financial risks.

Private investors have retained Team 54 to discover and analyze potential successful high-speed passenger rail corridors for development.

Overview and Deliverables

This document presents Team 54's approach for identifying priority city-pair corridors for high-speed passenger rail (HSR) development in the United States. To meet this objective, Team 54 analyzed travel patterns to forecast future demand and evaluated current transportation options to identify competitive advantages of HSR. Team 54 also developed and delivered a city-pair identification model and data-driven recommendations on priority high-speed rail corridors.

Analytics: ETL processes and comprehensive descriptive analytical reports providing historical travel demand information and comparison of transportation modes.

City-Pair Model: Analytic model detailing findings for identified optimal HSR corridors.

Recommendations: Data-driven recommendations providing HSR investors recommendations of city pairs. Future business recommendations are also included.

Data and Modeling

Five publicly available datasets, sourced primarily from government reports, were researched and obtained to build the city-pair profiles used in this report. Through extensive data cleaning, a high-quality combined city-pairs table was developed to represent the overall connection between cities.

The city pairs model developed by Team 54 is a strategic tool designed to identify the best potential routes for HSR development across the United States. By evaluating a range of factors-including population size, economic activity, traffic congestion, and travel demand-the model helps prioritize city pairs that offer the highest potential for successful HSR corridors.

Our **analysis considers both the individual strengths of each city**, such as its population and economic output, **and the relationship between city pairs**, including the distance between them and the demand for travel. A key aspect of the model is its ability to identify city pairs within the same geographic region, which naturally enhances the feasibility and success of HSR routes. The model ranks these city pairs, providing a clear roadmap for where private investment in HSR would likely see the greatest return.

To ensure a strategic approach to HSR development, Team 54 divided the United States into five key regions using node cluster mapping, each representing a cluster of cities with strong internal connections. Within each region, we identified the top city pairs that would most effectively serve as the foundation for an HSR network.

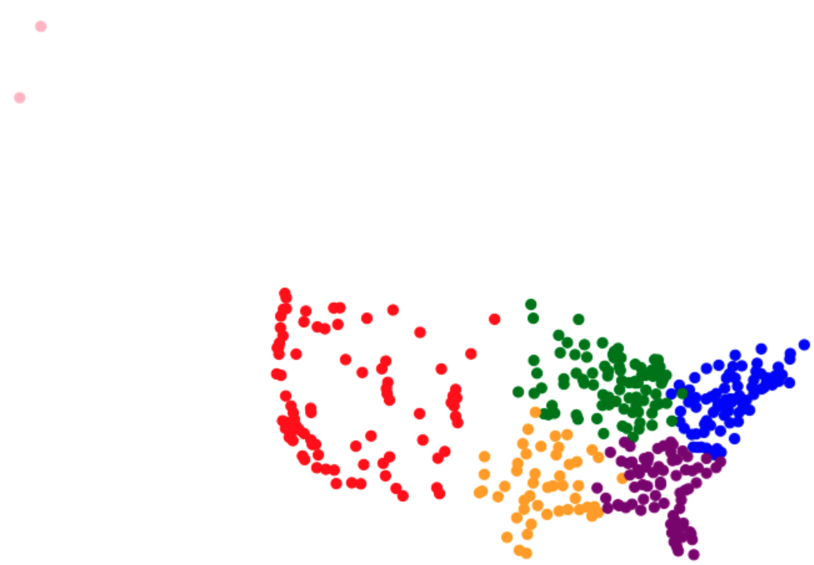
Key Insights

Team 54's analysis revealed useful insights into several areas of importance for the successful development of high-speed rail (HSR) corridors in the United States.

Regional Clusters

Team 54 used community detection, the process of grouping nodes in a network based on their internal connections, to identify clusters of cities for potential HSR corridor development. The results of this analysis show the optimal number of clusters to be six, five of which are viable in the continental United States. **This suggests that any initial development of HSR could potentially cluster corridors based on region.**

Figure 1. Optimal HSR community clusters in the United States

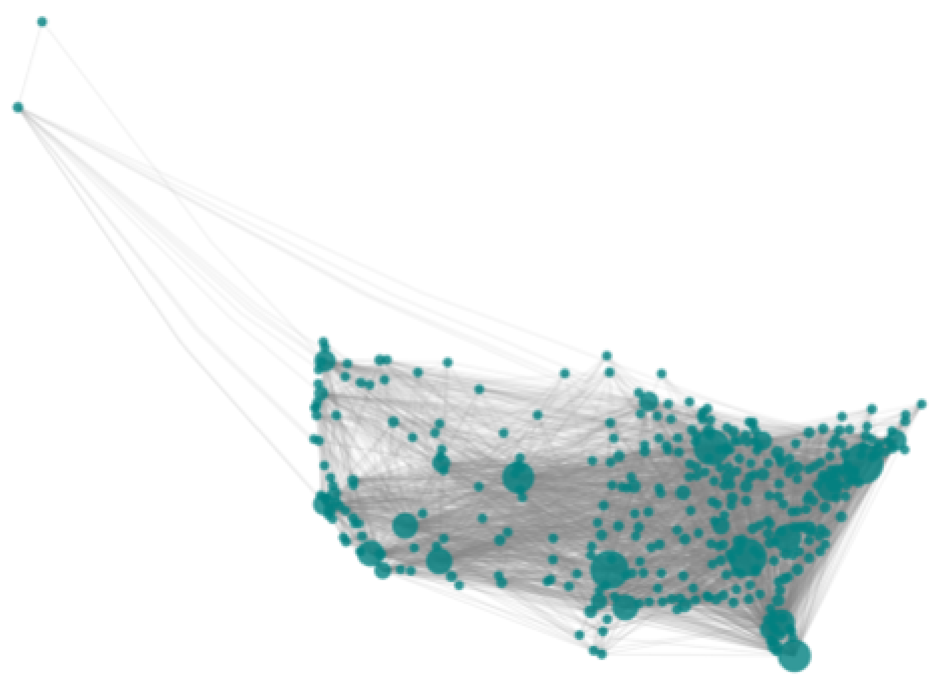


Travel Demand

Team 54 developed a demand score to capture the overall travel demand for each city in our dataset. The benefit of this score is that it represents the demand for a single city, rather than the demand for a particular route. This would be especially important for evaluating city pairs that do not have an existing flight connection but are both popular travel destinations.

This demand score played an important role in determining optimal city-pairs for HSR development.

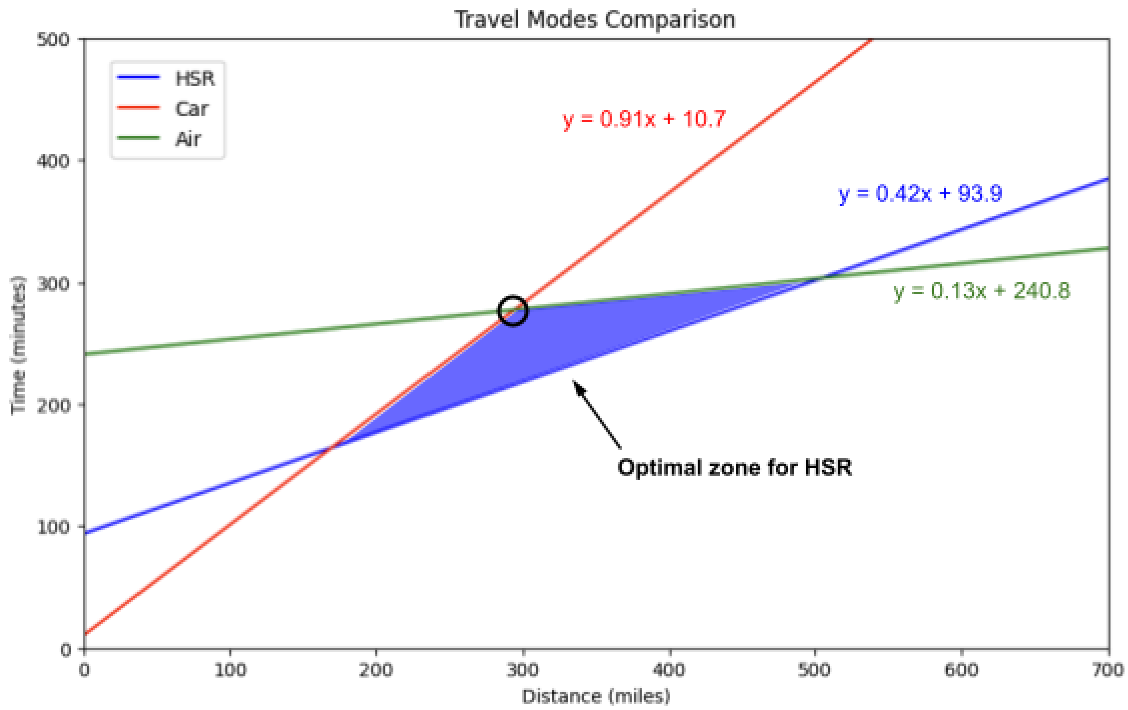
Figure 2. Map of cities with node size representing travel demand



Transportation Mode Comparison

The optimal distance for HSR to have a competitive advantage over other modes of travel is typically between 100 to 500 miles, taking into consideration that travelers prioritize trip time over distance and that travel time, including access and egress from airports or stations, plays a significant role in transportation mode decisions. **Team 54's analysis found an optimal HSR range of about 175 to 500, with the peak close to 300 miles.**

Figure 3. Travel Modes Comparison



Optimal City Pairs

The city pairs model emphasizes the importance of proximity, travel demand, population size, economic productivity, and congestion levels when identifying optimal city pairs for high-speed connections. The city pairs model identified strong candidates for high-speed rail connections based on the identified regional corridors.

In the Northeast, for example, city pairs like Boston-New York and New York-Washington, DC are prime candidates due to their high travel demand and economic significance. The Midwest centers around Chicago, connecting it with other major cities like Detroit and Minneapolis. The Southeast focuses on key routes linking major cities in Florida and Georgia, with Atlanta as a central hub. On the West Coast, routes between Los Angeles and cities like San Francisco and Las Vegas stand out due to their economic impact. Finally, the South Central region highlights Dallas as a key player in connecting Texan cities and beyond.

A sample of these findings is listed below and is fully discussed in the Analytic Approach section.

Table 1. Sample of Optimal HSR City Pairs

Origin	Destination	Score
Boston, MA	New York, NY	0.548827
New York, NY	Washington, DC	0.537625
Atlanta, GA	Miami, FL	0.531627
Dallas, TX	Houston, TX	0.476767
Los Angeles, CA	San Francisco, CA	0.463833
Las Vegas, NV	Los Angeles, CA	0.448542
Atlanta, GA	Orlando, FL	0.406607
Chicago, IL	Detroit, MI	0.373602
Chicago, IL	Minneapolis, MN	0.372956
Austin, TX	Dallas, TX	0.358635

Recommendations

Developing a high-speed rail (HSR) network in the United States is a complex project that requires a well rounded approach. HSR should be seen as part of a larger system that includes stations, maintenance, financing, and connections with existing transportation networks. Based on Team 54's analysis, we recommend the following strategies for successful HSR development:

Prioritize Key Regions: Focus on the Northeast Corridor, Midwest, Southeast, Western, and South Central regions, which offer the highest potential for initial HSR development. Key routes such as Boston-New York, Chicago-Detroit, Atlanta-Miami, Los Angeles-San Francisco, and Dallas-Houston should be prioritized due to their high potential for ridership and economic impact.

Enhance Regional Connectivity: Prioritize city pairs within regional corridors that fall within the optimal range for HSr, allowing them to effectively compete with air and auto travel. Strengthening connections within each region will reduce congestion and improve overall mobility.

Integrate Across Regions: Once regional HSR corridors are established, focus on connecting these regions to create a cohesive national network. Strategic connections identified through proximity analysis will ensure seamless travel options across the country.

Leverage Existing Infrastructure: Integrate new HSR lines with existing transportation networks, including airports, highways, and public transit, to create a seamless travel experience that enhances the appeal of HSR.

Monitor and Adapt to Demand: Continuously monitor travel demand and adjust services-such as train schedules and pricing- to meet passenger needs. This will ensure the network remains sustainable and maintain high ridership.

By following these recommendations, investors can strategically develop an HSR network that meet current transportation needs and supports long-term economic growth across the United States.

Introduction

Background

High-speed trains have successfully driven economic growth, created jobs, and brought communities together around the world. However, the benefits of high-speed rail have remained largely untapped in the United States, which remains reliant on auto and air travel. This began to change in 2009 when, as part of the American Recovery and Reinvestment Act (ARRA), the Federal Railroad Administration (FRA) was charged with distributing $8 billion for intercity and high-speed rail projects (Hagler & Todorovich, 2009). More recently, President Biden signed into law a $1.2 trillion infrastructure bill, which includes plans to invest $170 billion in improving and expanding the US rail network (Klein & Sullivan, 2021).

Lessons for the U.S. from Europe and Asia

In the years following World War II, Japan and European countries focused on rebuilding their railway systems. As a result of that investment, high-speed rail (HSR) first appeared on the world scene in 1964 with Japan's Shinkansen "Bullet Train." Since then, HSR has continued to expand in Europe and Asia, with extensive systems existing in China, France, Germany, Italy, Japan, and Spain.

During this same time period, the United States emphasized improvement to roadways and airports instead. Characteristics of the U.S., such as lower population density and an automobile-oriented culture, further reduced the competitiveness of passenger rail (Rutzen & Walton, 2011). However, high-speed trains offer the possibility of popularizing rail travel in the United States, as they offer major service improvements that make it more desirable to travelers (Rutzen & Walton, 2011).

Reviewing how Asian and European countries have developed and reorganized their railway systems can offer insights into what makes HSR successful. Each country has its own unique development path based on existing funding, infrastructure, and other commercial and operational aspects (Rutzen & Walton, 2011). For example, Japan's high-speed trains run on exclusive tracks, whereas Spain's tracks can be used by both high-speed and conventional trains.

However, commonalities also exist. For example, all Asian and European countries with major high-speed rail systems[1] had previous passenger rail systems in place (Rutzen & Walton, 2011). These national governments, along with the established agencies already in charge of rail infrastructure, have largely driven the decision to upgrade to HSR (Rutzen & Walton, 2011). In their paper *High Speed Rail: A Study of International Best Practices and Identification of Opportunities in the U.S.*, Rutzen and Walton (2011) write that strong government policies have played an important part in the success of high-speed rail over other modes of transport.

Their study also found that **identification of priority corridors is another key factor in the successful development of high-speed rail** (Rutzen & Walton, 2011). This report builds upon these findings by analyzing potential HSR corridors in the United States.

Business Case

In order for new U.S. high-speed rail projects to maintain public support, it is crucial that they are seen as successful. A major element of this success is choosing the right areas for high-speed rail development. Prioritizing corridors with the right population density, economic activity, and existing travel demand is essential for maximizing travel benefits and attracting strong ridership numbers (Hagler & Todorovich, 2009). Additionally, investing in corridors with the maximum potential to support high-speed rail systems reduces financial risks and increases the likelihood of long-term success and public support (Hagler & Todorovich, 2009).

Private investors have retained Team 54 to discover and analyze potential successful high-speed passenger rail corridors for development.

[1] China, France, Germany, Italy, Japan, Korea, The Netherlands, Portugal, Spain, Taiwan

Project Goals

To support the development of successful high-speed rail corridors, Team 54 analyzed travel demand and transportation modes and developed a model to identify the most beneficial city pairs in the United States for high-speed rail investment. Team 54 organized the work into three substantial goals that address all aspects of the project.

Model Travel Demand: Analyze travel patterns to forecast future demand for HSR.

Compare Transportation Mode: Assess scenarios in which HSR is a more efficient and sustainable option than air and road travel.

Identify Optimal City Pairs: Determine which city pairs in the U.S. have the highest potential for successful HSR implementation based on factors such as population density, economic activity, and travel demand.

Data

Overview

High-speed passenger rail does not currently exist in the United States. For this reason, Team 54 spent significant effort to research and understand the factors necessary to address this project successfully. Identifying available data sources that could be used to guide our analysis was essential. The focus of our data is on factors, such as metropolitan size, distance, and existing travel demand, that could be used to predict successful corridors for high-speed rail development.

Assumptions

Considering the significant lead time and inherent risks associated with high-speed rail investments, selected corridors must demonstrate strong potential for high passenger demand. In our analysis, we make several assumptions to support this notion. These assumptions, listed below, are supported by research from the America 2050 Prospectus (Hagler & Todorovich, 2009).

- City and metropolitan areas must have a population of at least 50,000.
- Distance between city pairs falls between 100-500 miles.
- City and metropolitan areas should have a high combined GDP.
- City and metropolitan areas should have a high level of auto congestion.

The rationale for each assumption is described below.

Metropolitan Size

Locating stations in major metropolitan areas helps ensure sufficient travel demand for HSR, as cities with larger populations tend to have higher potential ridership. Additionally, placing HSR stations in populous areas ensures that a significant portion of the population has easy access to the service. For this study, we use metropolitan statistical areas (MSA) delineated by the U.S. Office of Management and Budget (OMB), where each MSA must have at least 50,000 inhabitants (United States Census Bureau, n.d.).

Metropolitan statistical areas are composed of a core area with a substantial population nucleus and adjacent communities with strong economic and social ties to the core (United States Census Bureau, n.d.). Team 54 chose to work with MSAs rather than individual city data to obtain a more holistic view of potential ridership within an area. As of July 2023, there are 387 metropolitan statistical areas.

Distance

Numerous studies suggest an optimal distance between 100 to 500 miles for HSR to have a competitive advantage over other modes of travel. Distances below 100 miles are generally better suited for driving, whereas distances greater than 500 miles are best for flying. Team 54 validated this assumption by creating our own optimal distance model, while also accounting for travel time.

> The average traveler cares more about the time a trip takes than the distance traveled.

In this model, Team 54 chose to work with total travel time rather than pure distances between cities. Our assumption is that total travel time impacts a traveler's decision on transportation mode more so than pure distance. For example, air travel requires a significant time investment beyond the flight itself due to onerous security processes, long check-in times, and potential delays. Team 54 believes that travelers take these barriers to access into account when selecting their preferred travel mode. The results of this analysis can be found in the Analytic Approach - Compare Transportation Modes section.

GDP

The success of HSR systems will rely heavily on business travel to maintain ridership, with business travel being most prevalent in regions with highly productive economies. Studies indicate that travel—whether for business, personal, or leisure purposes—increases with increased income (Ewing, 2021). Gross Domestic Product (GDP) per capita serves as a comprehensive measure linking both economic productivity and personal income.

Congestion

HSR systems can significantly alleviate highway traffic congestion. As metropolitan congestion increases, intercity auto travel times become longer, making rail an increasingly attractive option. For example, studies conducted for the British railways suggested that a new HSR line would divert about 50% of the traffic from car and air travel (Rutzen & Walton, 2011). This diversion would, in turn, reduce congestion and delays in roads and airports (Rutzen & Walton, 2011).

Team 54's study utilizes the Texas Transportation Institute (TTI) Travel Time Index (TTI) for MSAs to identify metro areas with high levels of auto congestion. The Travel Time Index measures the ratio of travel time during peak traffic periods to travel time under free-flow conditions, highlighting areas where drivers experience the most significant delays (Texas A&M Transportation Institute, 2023).

Data Collection & Preprocessing

Five publicly available datasets, sourced primarily from government reports, were researched and obtained to build the city-pair profiles used in this report. Through extensive data cleaning, a high-quality combined city-pairs table was developed to represent the overall connection between cities.

Demographic Data	U.S. Census Bureau
Economic Data	Bureau of Economic Analysis (BEA)
Transportation Data	Department of Transportation (DOT) The Texas A&M Transportation Institute (TTI) Google Maps API
Geographic Data	Google Maps API

Metropolitan Statistical Area (MSA) Table

The MSA table is composed of data from four sources to represent the basic demographic, economic, and geographic characteristics of the cities considered in this analysis. For the purposes of this study, a city is represented by its MSA as defined in the U.S. Census Bureau's 2023 Metropolitan and Micropolitan Statistical Areas Population Totals report (United States Census Bureau, n.d.). This report was processed to extract and clean the viable MSAs and their 2023 population estimates.

To facilitate further enrichments, such as obtaining geo-coordinates via the Google Maps API, an additional field was created to represent the most populous city within each MSA.

Economic data, specifically Gross Domestic Product (GDP), was obtained from the Bureau of Economic Analysis (BEA) (Bureau of Economic Analysis, n.d.). Out of the 387 identified MSAs, 12 did not report GDP data and were therefore excluded from the table. Furthermore, the Time Travel Index was extracted from TTI's Urban Mobility Report to identify traffic congestion conditions for each MSA.

Table 2. Feature and Variable Description of MSA Table

Variable	Description
MetroArea	Metropolitan Statistical Area (2023)
MainCity	Most populous city within the MSA
Population	MSA population
Lat	Latitude
Lng	Longitude
GDP	GDP in thousands of dollars
TTI	Traffic congestion ratio

Drive Table

The drive table was constructed by querying the Google Maps Distance Matrix API for all combinations of Metropolitan Statistical Area (MSA) pairs. The API has a request limit of 100 elements, so queries were batched into groups of 10 origin and destination MSAs to improve efficiency. With 375 MSAs, this required nearly 1,500 query batches. The API returns the standard driving distance using the current road network and the average travel time for the route. Additional processing was performed on the results to clean the data and convert units.

Team 54 observed that, within the dataset, driving distances and durations were not always consistent between two cities. For example, the distance from Las Vegas to Los Angeles is 272 miles with an average travel time of 251 minutes, while the reverse trip from Los Angeles to Las Vegas is 270 miles with an average travel time of 243 minutes. To address this, the averages of the two trips were calculated to accurately reflect the city-pair driving distances and trip durations.

The resulting table consisted of over 70 thousand city pairs.

Table 3. Feature and Variable Description of Drive Table

Variable	Description
CityPair	MSA pair
Distance_meters	Averaged distance for route returned from API
Distance_miles	Calculated from Distance_meters
Duration_seconds	Averaged duration for route returned from API
Duration_minutes	Calculated from Duration_seconds

Flight Table

Flight data was obtained from the Department of Transportation's Bureau of Transportation Statistics (BTS) T-100 Domestic Segment dataset, which is a comprehensive collection of data on air traffic movements within the United States. This dataset provides detailed monthly information on the number of passengers, freight, and mail transported by certified air carriers operating domestic non-stop flights. For our analysis, Team 54 extracted data for the period of January to April 2024, amounting to over 38 MB and comprising more than 138,000 unique flight segments across 45 descriptive variables.

Extensive data cleaning was conducted to isolate relevant passenger flights for viable city pairs, resulting in a refined dataset of approximately 94,000 unique flight segments and 25 variables. Given that many large metropolitan areas are served by multiple airports and many MSAs consist of multiple cities, careful mapping of airports to the defined MSAs was crucial. To accomplish this, a list of world airport codes, as defined by the International Air Transport Association (IATA), was obtained from BTS. Regular expressions and string matching techniques were employed to compare cities within MSAs to the city and airport names associated with each airport code. These matching techniques revealed 18 MSAs that did not have an associated airport and were therefore excluded from the flight table.

After deriving the airport codes associated with the remaining MSAs, Team 54 used these codes to filter the T-100 dataset to obtain the flight segments between every MSA.

Table 4. Feature and Variable Description of Flight Table

Variable	Description
DEPARTURES_SCHEDULED	Number of scheduled flights
DEPARTURES_PERFORMED	Number of completed flights (not canceled)
SEATS	Number of seats available
PASSENGERS	Number of passengers on flights
DISTANCE	Distance between airports (miles)
RAMP_TO_RAMP	Ramp to Ramp Time (minutes) (cumulative)
AIR_TIME	Airborne Time (minutes) (cumulative)
UNIQUE_CARRIER	Unique Carrier Code
AIRLINE_ID	Identification number assigned by DOT to identify a unique carrier
UNIQUE_CARRIER_NAME	Unique Carrier Name
REGION	Carrier's Operation Region
CARRIER_GROUP_NEW	Carrier Group Code (i.e. Foreign carrier, Large Regional, etc.)
ORIGIN_CITY_MARKET_ID	Identification number assigned by DOT to identify a city market.
ORIGIN	IATA airport code for origin
ORIGIN_CITY_NAME	Origin city, state
DEST_CITY_MARKET_ID	Identification number assigned by DOT to identify a city market
DEST	IATA airport code for destination
DEST_CITY_NAME	Destination city, state

AIRCRAFT_GROUP	Group of aircraft
AIRCRAFT_TYPE	Type of aircraft
YEAR	2024
MONTH	Month (1-4)
DISTANCE_GROUP	Distances grouped in 500 mile increments
CLASS	Type of flight (i.e. scheduled passenger, unscheduled passenger)
CityPair	Calculated during processing to correspond to MSA

Data Post Processing & Transformations

After the initial data tables were created, additional processing was required to prepare the data for modeling. Post-processing and transformations included creating a combined city-pairs table and loading the table into a graph database, Neo4j, to compute features used in the model.

Combined City Pairs Table

Team 54 recognized the need to create a combined city-pairs table in order to better represent the overall connection between cities. This was achieved by merging the drive and flight tables.

First, the flight table was grouped by city pairs, and all variables were aggregated by summing the number of passengers for all flight segments for a city pair. The resulting dataset was then refined to include only those variables that were relevant for the analysis, such as distance, duration, flight frequency, and passenger volume. To ensure all routes were accounted for, including those without flights between a city pair, Team 54 imputed missing values. Geodesic distance (the shortest path between two points on a curved surface) was calculated for missing flight distance values. All other null values were set to zero. Lastly, all variables were normalized using the difference between the minimum and maximum values.

Table 5. Feature and Variable Description of Combined City Pair Table

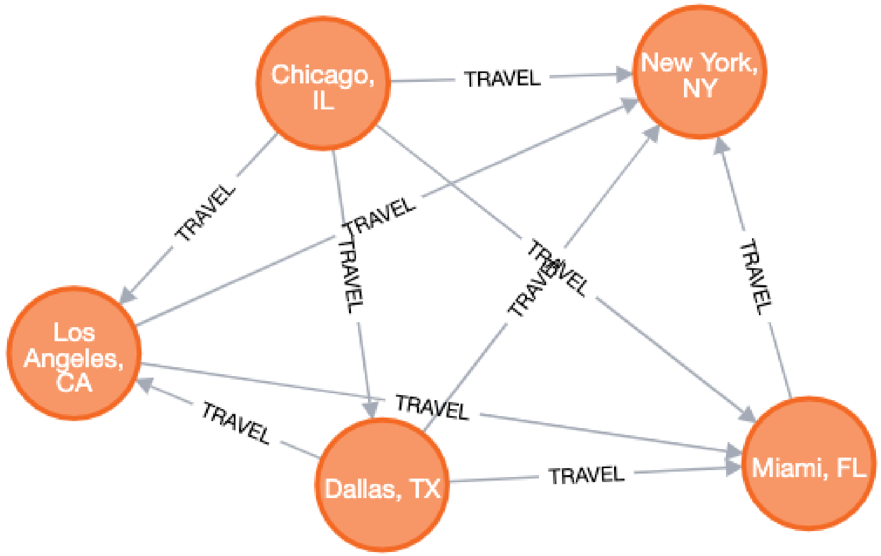
Variable	Description
drive_distance	Google Maps API avg distance (miles)
drive_duration	Google Maps API avg duration (minutes)
flight_distance	Average distance (miles) from flight table. Impute null values with geodesic distance
flight_duration_ramp	Average duration (minutes) ramp_to_ramp
flight_duration_air	Average duration (minutes) in the air
flight_duration_total	Average duration (minutes) ramp_to_ramp + 200 min time penalty
num_passengers	Cumulative number of passengers that traveled during the 4 months
total_seats	Cumulative number of seats available on flights
flight_frequency	Cumulative number of flights during the 4 months
scheduled	Cumulative number of flights scheduled
num_carriers	Cumulative number of airlines operating between cities

Graph Database

To further analyze the complex relationships between the cities and compute additional features using graph data science techniques, the data was loaded into Neo4j, one of the leading graph databases. Neo4j uses the Cypher query language, which is a declarative query language specifically designed for graph databases. Cypher simplifies the process of expressing complex graph patterns and extracting meaningful insights from interconnected data.

Using Python's Neo4j driver, Team 54 successfully modeled **MSA nodes** and **route relationships** in the graph. The MSA nodes represent each city as a node with properties from the MSA dataframe, such as population, GDP, traffic congestion Travel Time Index (TTI), latitude, and longitude. Route relationships connect city nodes using relationships that represent city-pair travel information, such as drive distance and duration, flight distance and duration, and passenger counts.

Figure 4: Sample visual of MSA nodes and route relationships



In addition, an efficiency score was calculated for each city pair. The efficiency score is a composite metric that looks at the route relationship for each city pair and aims to maximize passenger volume while minimizing distance. This score was then added to the Combined City Pair Table so that it could be easily accessed for the Optimal City Pair model.

All variables were normalized using min/max normalization to scale the data to a range of [0,1] based on the minimum and maximum values in the dataset. This ensured a consistent scale across all efficiency scores. Normalization was completed in Python and then added as a property to the relationship in Neo4j.

Table 6. Top 10 cities based on efficiency score

City1	City2	PassengerVolume	DriveDistance	EfficiencyScore
Boston, MA	New York, NY	895039.0	215.741565	5.206261
Colorado Springs, CO	Denver, CO	253850.0	70.696796	5.062664
Las Vegas, NV	Los Angeles, CA	1051033.0	270.797521	4.818183
Dallas, TX	Houston, TX	869777.0	239.181232	4.539664
Los Angeles, CA	San Francisco, CA	1213766.0	382.824809	3.887872
Miami, FL	New York, NY	4152861.0	1285.911440	3.879758
Austin, TX	Dallas, TX	541007.0	195.118261	3.724418
Atlanta, GA	Miami, FL	2042554.0	664.196511	3.499368
Charlotte, NC	Durham, NC	364155.0	144.142227	3.257093
New York, NY	Washington, DC	568869.0	226.807871	3.139336

Graph Data Science Algorithms

Neo4j has a Graph Data Science Library that includes a comprehensive suite of algorithms for centrality, community detection, similarity, and pathfinding.

Centrality

Centrality is a social network analysis metric that measures the degree to which a person, organization, or in this case, a city, is central to a network. Centrality measures were used to help identify key city hubs that can maximize network connectivity if connected by high-speed rail.

Degree centrality is a measure of the number of connections (edges) a node has in a network. For example, cities that are directly connected to many others are considered more central. In a weighted context, degree centrality can be interpreted as the sum of the weights of all edges connected to the node. Team 54 assessed weighted degree centrality as a viable method to quantify and leverage existing travel data more effectively. The idea is that a MSA with a high degree centrality based on travel demand will likely also have a high demand in new transport modalities like HSR. Further details on the implementation and results of this approach can be found in the Model Travel Demand section.

Closeness centrality is a measure of the closeness or distance of one node to others in a network. More central nodes can quickly reach all other nodes in the network. In the case of HSR, this allows us to identify cities that are central in terms of access to all other cities in the network. The closeness centrality algorithm in Neo4j does not support weights, making it unusable for this project. However, Team 54 developed a similar version of the Neo4j algorithm that includes weights by aggregating proximity scores for each city based on inverse distances.

Calculating inverse distances approximates the idea of closeness centrality with weights. A higher proximity score for a city indicates that it is "closer" to other cities based on the cumulative inverse distances. Summing the inverse distances (or proximity scores) from a city to all others provides a measure akin to closeness centrality, reflecting the city's accessibility or central location relative to the rest of the network.

The cities listed in these results are unlikely to be the first to receive HSR due to other factors such as lower population, GDP, and overall travel demand. However, their high proximity scores indicate that these cities could play an important future role in tying together various regional HSR corridors, as HSR expands across the United States.

Table 7. Top 10 cities based on proximity score

City	TotalProximityScore
South Bend, IN	0.856574
Niles, MI	0.844326
Elkhart, IN	0.809337
Gettysburg, PA	0.806830
Harrisburg, PA	0.795045
York, PA	0.791775
Chambersburg, PA	0.784695
Lancaster, PA	0.778519
Hagerstown, MD	0.776615
Indianapolis, IN	0.771204

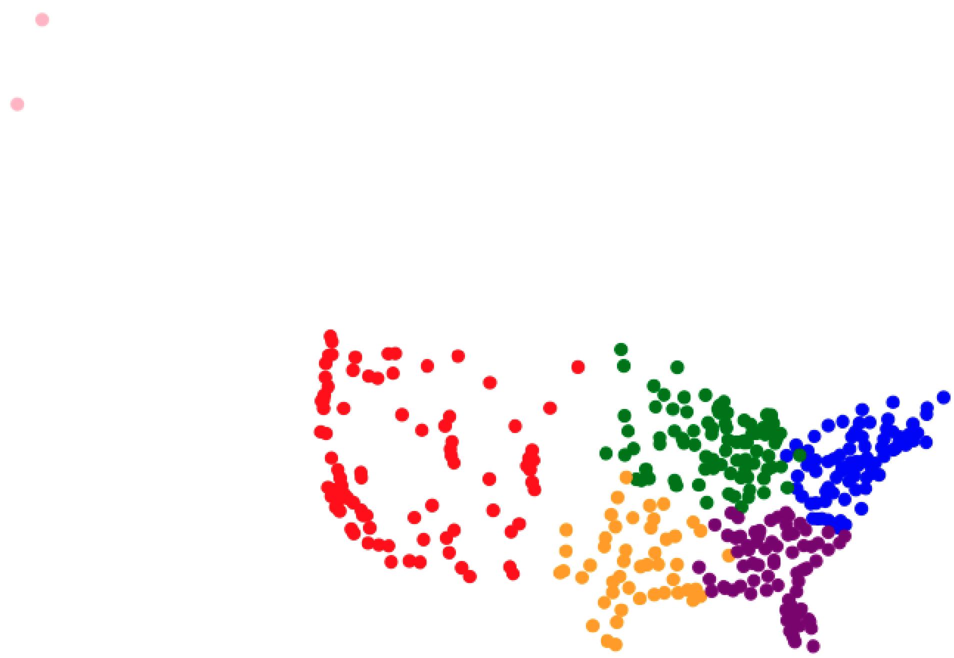
Community Detection

Community Detection is the process of grouping nodes in a network based on their internal connections. For this project, it was used to identify clusters of cities for potential HSR corridor development. Team 54 experimented with different algorithms, including K-means clustering using geo-coordinates and Leiden and Louvain algorithms using weighted distance measures. The best results came from the Leiden algorithm.

The Leiden algorithm is a hierarchical clustering algorithm that separates nodes into disjoint communities to maximize a modularity score for each community. The number of communities is determined automatically based on the network structure. Additionally, unlike K-means, the Leiden algorithm allows a relationship property to be used as weight. Team 54 used city-pair distance as the weight.

A Gaussian (bell curve) function was used to give the highest weight to distances close to the optimal travel distance for HSR: 300 miles (the optimal travel distance is further discussed in the Compare Transportation Modes section). These weights gradually decreased as city-pair distances moved away from the optimal 300-mile value. The results, as displayed in Figure 5, show the optimal number of communities/clusters to be six, five of which are viable in the continental United States. This suggests that any initial development of HSR could potentially cluster corridors based on region.

Figure 5: Optimal HSR community clusters in the United States



Analytic Approach

Model Travel Demand

In the absence of HSR in the US, Team 54 has relied on historical flight data to gauge travel demand between city pairs. To comprehensively assess which metrics best represent travel demand, we considered several variables including passenger numbers, flight frequencies, and carrier counts between city pairs. We also observed the impact that distance appears to have on each. Our analysis shows that these metrics, particularly the number of passengers and the frequency of flights, are crucial indicators of demand, regardless of the distances involved. This observation is supported by correlation analyses which suggest that distance does not significantly impact the demand for travel between city pairs. The accompanying graphs in this section display the top 20 city pairs for each metric.

Figure 6: Top 20 City Pairs by Number of Passengers and Distance

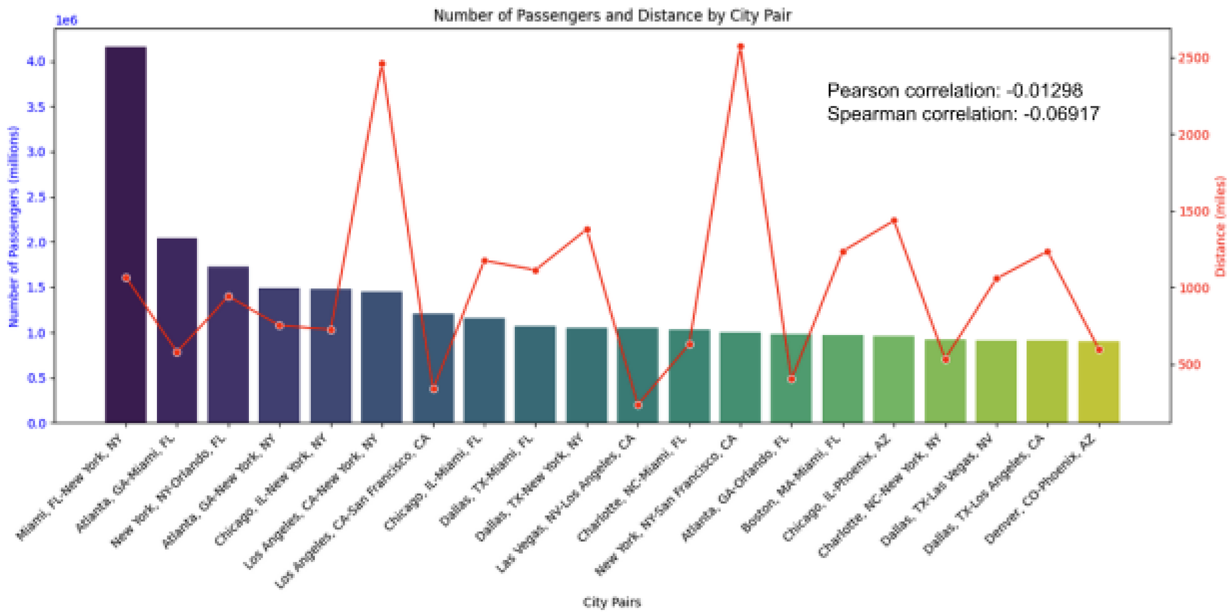


Figure 7: Top 20 City Pairs by Scheduled Flights and Distance

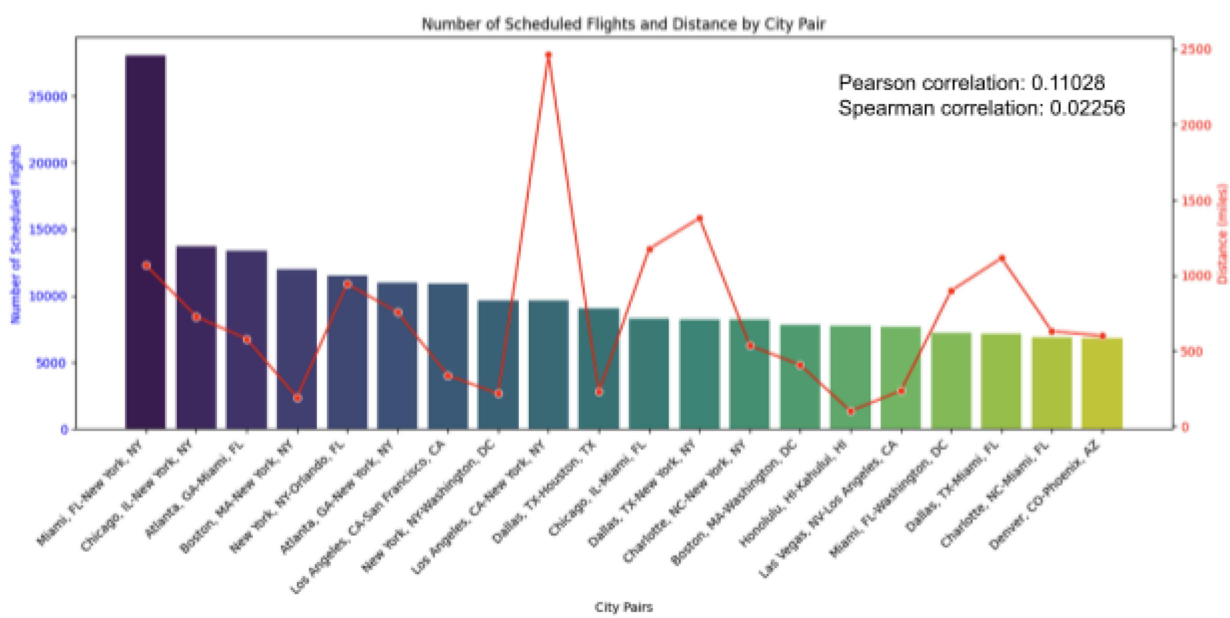
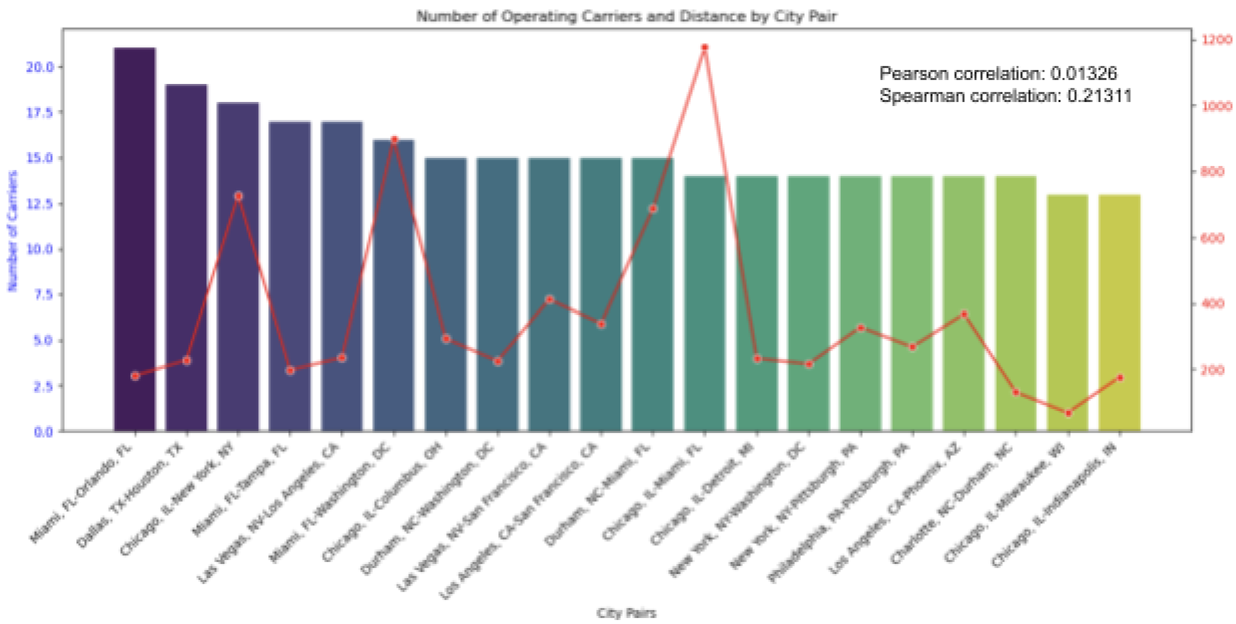


Figure 8: Top 20 City Pairs by Operating Carrier and Distance



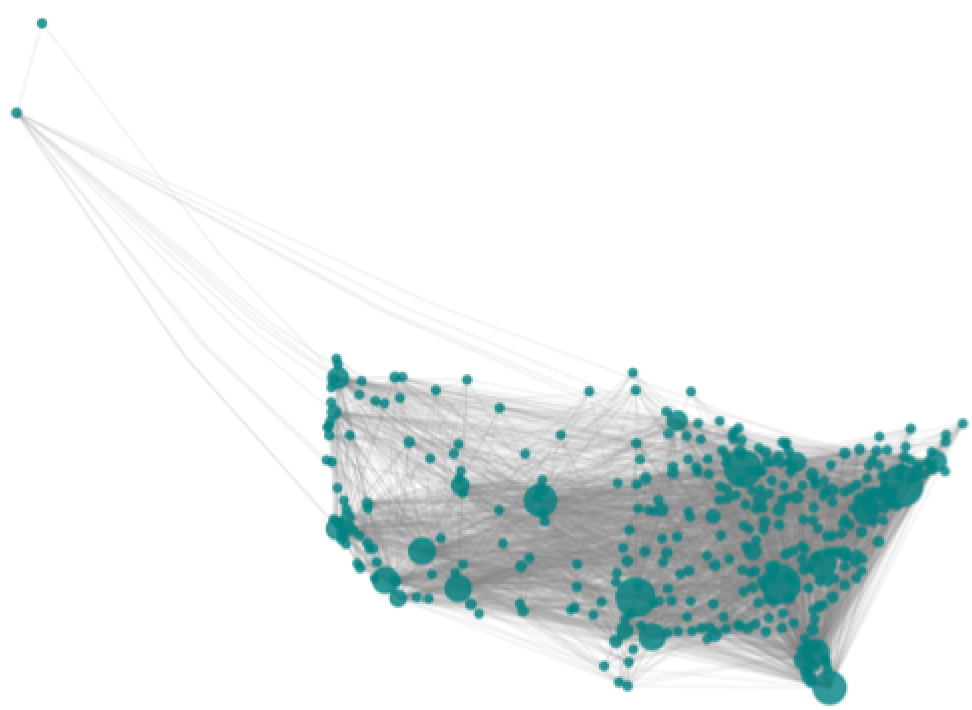
Based on this analysis to evaluate variables implicating travel demand, Team 54 chose to prioritize the number of passengers and flight frequency for our weights. This was accomplished by defining a custom metric: *demand_weight = num_passengers * 0.7 + flight_frequency * 0.3*. Before applying the weight, all values were first normalized using min/max scaling. This weight was then utilized to calculate degree centrality for each MSA.

Table 8 displays the centrality score for the top ten cities based on the weighted degree centrality metric. This centrality score reflects the total volume of these weights, effectively capturing the overall travel demand for the city. Using this score, Team 54 was able to represent the demand for a single city, rather than the demand for a particular route. This would be especially important for evaluating city pairs that do not have an existing flight connection but are both popular travel destinations. As a hypothetical example, if no flights existed from New York to Dallas, we could still identify this route as a good candidate for HSR because both have a high degree centrality score. Figure 6 displays a visualization of the MSA's with the node size indicating the travel demand score.

Table 8. Top 10 cities based on centrality score

MSA	DemandScore
New York, NY	7.544421
Dallas, TX	6.976007
Atlanta, GA	6.885777
Chicago, IL	6.449557
Miami, FL	5.577782
Denver, CO	5.413152
Charlotte, NC	4.411064
Orlando, FL	4.119093
Phoenix, AZ	4.045974
Los Angeles, CA	4.017987

Figure 9. Map of cities with node size representing travel demand



Compare Transportation Modes

As discussed in the Assumptions section, Team 54 validated the common assumption that the optimal distance for HSR ranges from 100 to 500 miles. To achieve this, we incorporated travel time into our analysis, emphasizing that door-to-door travel time impacts travelers' decisions more significantly than the distance traveled. This travel time includes estimated fixed time costs for access and egress from airports or stations (such as travel to/from, security checks, boarding, and disembarking) along with the actual onboard time. Figure 10 illustrates the estimated time penalties for each transportation mode.

Figure 10. Time penalties by transportation mode

Access To Main Transport		Main Transportation Mode	Egress From Main Transport		Total Added Time
0	0	CAR	0	0	0
30 min	2 h[1]	AIR	20 min	30 min	3 h 20 min
15 min	30 min[2]	HSR	10 min	15 min	1 h 10 min

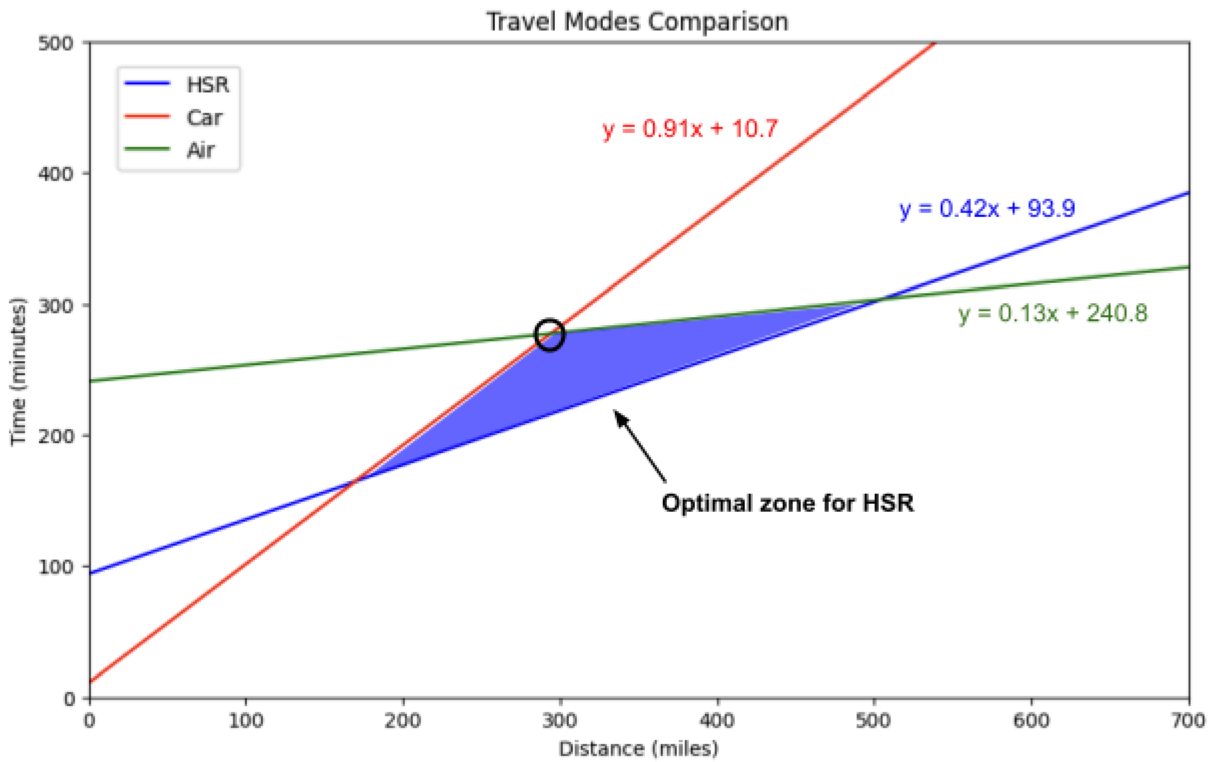
Team 54 developed base linear models for each mode of transportation by performing linear regression on random samples (n=25) of viable city pairs. For car travel, we utilized trip distance and time data from the Google Maps Distance Matrix API. Air travel data was sourced from the Department of Transportation, and the average air trip distance and time (from ramp to ramp) were computed for each city-pair. Since HSR data is not available for

[1] Per airline recommendation for domestic flights
[2] Per Amtrak recommendation

the US, we manually collected timetable data for 15 international HSR routes. A table detailing these international HSR routes can be found in the appendix.

After establishing the base models, we added the fixed time penalties to each mode. Graphing these results demonstrated an optimal HSR range of approximately 175 to 500 miles, with the peak efficiency close to 300 miles.

Figure 11. Travel Modes Comparison



City Pairs Model

The city pairs model is a sophisticated analytical tool designed to identify optimal city pairs for high-speed rail (HSR) connections by evaluating multiple critical factors. These factors were carefully selected to assess the viability and potential success of HSR routes, considering both the characteristics of individual metropolitan areas (MSAs) and the relationships between city pairs. The model integrates a combination of demographic, economic, and logistical criteria to ensure that the identified city pairs are not only feasible but also offer significant benefits in terms of ridership, economic impact, and congestion relief.

MSA (Metropolitan Statistical Area) Contributions

The scoring model considers the unique characteristics of each city involved in a city pair, with the following factors being important:

- **Population:** The population of an MSA is a critical factor, as large populations typically correlate with higher potential demand for HSR services. The population score was normalized and weighted at 0.3 in the model, reflecting its moderate importance in determining the viability of a route.
- **GDP:** An MSA's economic productivity, represented by its GDP, indicates its financial and infrastructural capacity to support HSR development. Higher GDPs suggest stronger economic activity and greater potential for both business and leisure travel. This factor was weighted at 0.2, emphasizing its role in ensuring that selected routes can sustain high-speed rail operations.
- **Travel Time Index (TTI):** The Travel Time Index measures the severity of traffic delays and measures congestion levels in an MSA. MSAs with higher congestion scores are prioritized, as HSR offers a faster and more reliable alternative to road travel, potentially alleviating congestion and reducing environmental impacts. This factor was weighted at 0.1, indicating its role in enhancing the appeal of HSR in congested regions.
- **Travel Demand:** Perhaps the more critical factor, travel demand reflects the overall demand for transportation within an MSA. It captures the existing need for efficient, high-capacity travel solutions. The demand score was normalized and given the highest weight among MSA contributions at 0.4, underscoring its importance in predicting the success of an HSR route.

These MSA-specific factors were aggregated for both cities in a pair, with each city's contribution forming half of the MSA score. The total MSA contribution accounts for 50% of the final score for each city pair.

Relationship (City Pair) Contributions

The relationship between the cities in a pair is also a critical component, evaluated with the following criteria:

- **Distance:** The physical distance between city pairs is a crucial factor in determining the competitiveness and practicality of HSR relative to other modes of transport, such as air and car travel. This criterion was evaluated using a Gaussian distribution that emphasizes the optimal distance for HSR, which we identified as being around 300 miles. This distance is considered ideal because it allows HSR to provide a faster and more convenient travel option, particularly when compared to the time required for air travel, which includes airport security and boarding processes. The Gaussian distribution used in the model assigns higher scores to distances close to this optimal range, ensuring that routes are selected where HSR offers a distinct advantage. This criterion carries a weight of 0.4, reflecting its significant impact on the overall viability of the HSR route.
- **Efficiency:** The efficiency score in this context specifically measures how effectively travel demand can be met relative to the distance between city pairs. It is calculated by considering the travel demand divided by the distance, essentially evaluating whether the potential ridership justifies the length of the route. A higher score indicates that a route has a strong demand relative to its distance, making it a more efficient option for HSR development. This criterion was normalized and carries a weight of 0.1 in the model. Although it plays a role in ensuring that routes are practical and resource-efficient, it is weighted less heavily compared to other factors such as overall demand, as it serves more as a balancing measure to ensure routes are not excessively long relative to their demand.
- **City Pair Demand:** Demand for transportation between the two cities in a pair is the most critical factor in the relationship contributions. This score reflects the specific demand for travel between the cities, based on historical data, particularly from flight travel, which provides a strong indication of existing travel patterns and the potential ridership for an HSR route. High demand between two cities suggests that an HSR service would be well-utilized, justifying the investment and operational costs associated with developing the route. Due to its importance in predicting the success

of an HSR route, this criterion was given the highest weight in the relationship contributions at 0.5, emphasizing that strong demand between city pairs is essential for the viability of the proposed routes.

Cluster-Based Scoring

This factor considers whether two cities in a pair belong to the same regional cluster, as defined in the Graph Data Science Algorithms Section.

- **Cluster:** If both cities in a pair belong to the same geographic cluster, the model added a boost of 0.15 to the score. This boost recognizes that connecting cities within the same cluster is generally more desirable for HSR development due to existing synergies surrounding an interconnected sector and shorter distances.

Total Score Calculation

The total score for each city pair was calculated as follows:

- **City Contributions:** The contributions from both cities (City 1 and City 2) were calculated by combining their MSA contributions (population, GDP, TTI, demand). The combined weight of these city-specific factors was split evenly between the two cities, each contributing 25% to the total score.
- **Pair Contributions:** The pair-specific factors (distance, efficiency, and city pair demand) together contribute 50% of the total score.
- **Cluster Boost:** An additional boost is added if both cities belong to the same cluster, further enhancing the total score.

Normalization and Output

After calculating the total score, the model normalized these scores by dividing by the maximum score, ensuring that the highest score scales to a standard value. This normalization process allowed for easier comparison and ranking of city pairs.

The model then outputted a sorted list of city pairs, ranked by their normalized scores. **The top-ranking city pairs are those that offer the greatest potential for successful HSR development, based on their characteristics.**

Table 9. Top 25 HSR City Pairs Overall

City1	City2	Total_Score
Miami, FL	New York, NY	0.652469
Boston, MA	New York, NY	0.587351
Atlanta, GA	Miami, FL	0.572455
New York, NY	Washington, DC	0.570617
New York, NY	Portland, ME	0.565233
Los Angeles, CA	San Francisco, CA	0.552525
Burlington, VT	New York, NY	0.548907
Las Vegas, NV	Los Angeles, CA	0.546564
New York, NY	Richmond, VA	0.540415
Buffalo, NY	New York, NY	0.529744
Charlottesville, VA	New York, NY	0.528701
Dallas, TX	Houston, TX	0.521576
Chicago, IL	Louisville, KY	0.519867
Johnstown, PA	New York, NY	0.519308
New York, NY	Virginia Beach, VA	0.518518
Chicago, IL	New York, NY	0.515339
Lewiston, ME	New York, NY	0.510243
New York, NY	Pittsburgh, PA	0.509683
Dallas, TX	Little Rock, AR	0.508377
Boston, MA	Philadelphia, PA	0.505124
Harrisonburg, VA	New York, NY	0.504265

Los Angeles, CA	San Jose, CA	0.502676
Las Vegas, NV	Phoenix, AZ	0.500251
Los Angeles, CA	Modesto, CA	0.498517
Chicago, IL	St. Louis, MO	0.495997

City Pairs by Region defined by Data Mapping

The city pairs model with regional scoring seeks to identify the best city pairs for launching high-speed rail (HSR) routes across different regions of the United States. This approach divides the country into five distinct clusters, or regions, and evaluates potential city pairs within each region using a set of weighted criteria. The primary goal is to identify the most effective city pairs to initiate an HSR network within each region, considering local factors that influence the viability and success of these routes.

The clustering of regions was defined through an analysis using Neo4j's Graph Data Science Algorithms, specifically focusing on community detection methods. The process involved utilizing centrality metrics and the Leiden algorithm to group cities based on their internal connections and proximity. The Leiden algorithm was particularly effective, as it allowed the use of weighted distance measures, which were crucial for identifying optimal clusters for HSR development. The results of this clustering, as shown in Figure 5: Optimal HSR Community Clusters in the United States, indicate five viable regions within the continental U.S. These clusters are designed to maximize connectivity within each region, making them ideal starting points for an HSR network.

In each region, the model evaluated city pairs based on their population, GDP, traffic congestion (TTI), and travel demand, as well as the relationship between city pairs, including distance, efficiency, and specific city pair demand. By scoring and ranking city pairs within these defined regions, the model helps prioritize routes that would be most effective in initiating HSR networks across the country.

Specificities of the Regional Model

Regional Scoring: The regional model included a regional scoring system where city pairs are evaluated within their respective regions or clusters. Each region was analyzed independently to identify the top city pairs, which are expected to serve as the initial connections for an HSR network within that region. The model evaluated the potential effectiveness of these connections using a combination of unique contributions from each metropolitan statistical area (MSA) and relationship (city pair) contributions.

Weights Assigned to Criteria: The scoring for each region was based on specific weighted criteria, divided into two main categories:

- MSA Unique Contributions:
    - Population Weight: 0.2
    - GDP Weight: 0.2
    - Traffic Congestion (TTI) Weight: 0.1
    - Travel Demand Score: 0.5
- Relationship Contributions (City Pair Contributions):
    - Distance Weight: 0.2
    - Efficiency Weight: 0.2
    - City Pair Demand Weight: 0.6

The high emphasis on travel demand (both regionally and for specific city pairs) reflects the model's focus on selecting routes that are likely to have high ridership and therefore greater chances of success.

Region 1 - Northeast Corridor

The Northeast Corridor is a densely populated region with significant existing demand for intercity travel. The top city pairs reflect this, with major cities like New York, Boston, and Washington, DC featuring prominently. These routes are already well-served by existing rail services, and the addition of HSR could further enhance connectivity in this economically vital region. Notably, New York, NY consistently achieved some of the highest criteria scores when normalized across all U.S. cities, reinforcing its role as a pivotal hub in the region.

Table 10. Top 5 HSR City Pairs in Northeast Corridor

City1	City2	Regional_Score
Boston, MA	New York, NY	0.548827
New York, NY	Washington, DC	0.537625
New York, NY	Philadelphia, PA	0.429902
New York, NY	Pittsburgh, PA	0.409905
Durham, NC	New York, NY	0.406129

Region 2 - Midwest Corridor

Chicago is the central hub in this region, connecting to various key cities in the Midwest. The high rankings of these city pairs suggest that Chicago's connectivity with other Midwestern cities is crucial for the success of HSR in this region. The focus on routes that connect major economic centers within relatively short distances supports the idea of HSR as a viable alternative to air and car travel in the Midwest.

Table 11. Top 5 HSR City Pairs in Midwest Corridor

City1	City2	Regional_Score
Chicago, IL	Detroit, MI	0.373602
Chicago, IL	Minneapolis, MN	0.372956
Chicago, IL	St. Louis, MO	0.341534
Chicago, IL	Cincinnati, OH	0.322334
Chicago, IL	Louisville, KY	0.321587

Region 3 - Southeast Corridor

The Southeast region focuses heavily on connections between major cities in Florida and Georgia. Atlanta, as a major economic hub, is a key player, with routes connecting it to other significant Southeastern cities like Miami and Orlando. The high rankings of these city pairs indicate strong demand and the potential for these routes to anchor HSR development in the Southeast.

Table 12. Top 5 HSR City Pairs in Southeast Corridor

City1	City2	Regional_Score
Atlanta, GA	Miami, FL	0.531627
Atlanta, GA	Orlando, FL	0.406607
Charlotte, NC	Miami, FL	0.375754
Atlanta, GA	Charlotte, NC	0.367011
Miami, FL	Orlando, FL	0.355756

Region 4 - Western Corridor

This region highlights key city pairs that connect major population centers on the West Coast with large inland cities. The connections between Los Angeles and other major cities like San Francisco and Las Vegas are particularly strong due to high demand and economic activity, making them ideal candidates for early HSR development.

Table 13. Top 5 HSR City Pairs in Western Corridor

City1	City2	Regional_Score
Los Angeles, CA	San Francisco, CA	0.463833
Las Vegas, NV	Los Angeles, CA	0.448542
Denver, CO	Los Angeles, CA	0.404602
Los Angeles, CA	Phoenix, AZ	0.392027
Los Angeles, CA	Seattle, WA	0.353088

Region 5 - South Central Corridor

Texas dominates this region, with Dallas serving as a major hub for multiple routes. The strong economic ties and high travel demand between these Texan cities, along with connections to neighboring states, suggest that these routes could be highly successful in an HSR network. The model shows that there is significant potential for regional HSR development focused on Dallas as a central point.

Table 14. Top 5 HSR City Pairs in South Central Corridor

City1	City2	Regional_Score
Dallas, TX	Houston, TX	0.476767
Austin, TX	Dallas, TX	0.358635
Dallas, TX	San Antonio, TX	0.348523
Dallas, TX	Little Rock, AR	0.312061
Dallas, TX	New Orleans, LA	0.301269

Conclusion

Team 54’s comprehensive analysis of potential high-speed rail (HSR) corridors across the United States has identified several high-priority city pairs that are well-positioned for successful HSR development. The city pairs model we developed evaluates critical factors such as population, economic activity, travel demand, and transportation efficiency to determine which routes offer the greatest potential for ridership and economic benefits. By focusing on city pairs within defined regions, we have provided a strategic roadmap for HSR investment that maximizes connectivity and regional growth.

Our analysis revealed that clustering MSAs based on geographical proximity was a powerful strategy for enhancing the viability of HSR routes. By incorporating community detection techniques, we identified regional clusters where intra-cluster city pairs show the highest potential for successful HSR implementation. Additionally, the inclusion of normalized economics and demographic variables allowed us to capture the broader socio-economic impact of HSR, ensuring that selected routes are not only feasible but strategically beneficial.

It is important to note that while this model is robust, it is not exhaustive. Certain factors such as transit system integration, regional topography, and detailed environmental impact assessments were not incorporated into this model. These additional factors may be worthy of future research to provide an even more comprehensive analysis of potential HSR routes.

The results of this study highlight several city pairs that stand out as strong candidates for high-speed rail connections. Notably, each region appeared to have a “hub” that served as a central node in the HSR network. These hubs arose based on their significant population density, economic influence and existing travel demand, as well as an ideal distance score.

Northeast Corridor: New York, NY emerged as the hub in this region, serving as a major connection point for cities like Boston, MA and Washington, DC.

Midwest Corridor: Chicago, IL stands out as the primary hub for the midwest region connecting other key cities such as Detroit, MI and St. Louis, MO.

Southeast Corridor: Atlanta, GA and Miami, FL arose as the primary cities for this region, connecting to other significant regional cities such as Charlotte, NC and Orlando, FL.

Western Corridor: Los Angeles, CA stands out as the key city in the western region, connecting to other notable cities including San Francisco, CA and Las Vegas, NV.

South Central Corridor: Dallas, TX is the focal point in the south central region, linking other major urban areas such as Austin, TX and New Orleans, LA.

Recommendations

Developing a high-speed rail (HSR) network across the United States is a complex and ambitious endeavor that requires a holistic approach. HSR should not be viewed as a standalone infrastructure project but as a part of a broader system that includes stations, maintenance, financing, and integration with existing transportation networks. Based on Team 54's extensive analysis of travel demand, transportation modes, and the identification of optimal city pairs, we propose the following recommendations for the successful development of HSR in the U.S.

1. **Prioritize Investment in Key Regions:** Our analysis indicates that the Northeast Corridor, Midwest, Southeast, Western, and South Central regions offer the highest potential for initial HSR development. These regions should be the primary focus for investment, with particular attention given to the top ranking city pairs identified in our model, such as Boston-New York, Chicago-Detroit, Atlanta-Miami, Los Angeles-San Francisco, and Dallas-Houston. These routes are expected to generate high ridership and significant economic benefits, making them ideal candidates for early HSR projects.
2. **Focus on Regional Connectivity:** To maximize the effectiveness of HSR, it is essential to prioritize regional corridors that keep city pairs within the optimal competitive range for HSR, allowing them to effectively compete with air and auto travel. By focusing on regional connectivity, we can enhance the interconnectivity of cities within each cluster, reducing congestion on roads and at airports. As the HSR network matures, attention should be directed towards connecting secondary cities within each region, thus promoting regional economic development and improving overall mobility.
3. **Integrate Regional HSR Corridors:** As regional HSR corridors are established, the next phase should focus on integrating these corridors across different regions. The use of closeness centrality measures, such as the proximity score from our Graph Data Science Algorithm analysis, can help identify the most strategic connections between regional HSR networks. This approach will ensure that the U.S. develops a cohesive and interconnected HSR system, providing seamless travel options across the country.

4. **Leverage Existing Transportation Infrastructure:** To enhance the feasibility and cost-effectiveness of HSR projects, new rail lines should be integrated with existing transportation infrastructure. This includes ensuring HSR stations are connected with major airports, highways, and public transit systems for optimal fit. By doing so, we can create a seamless travel experience for passengers, reducing overall travel time and increasing the appeal of HSR as a competitive alternative to other modes of transportation.
5. **Monitor and Adjust Based on Demand:** As HSR routes are developed, it is vital to continuously monitor travel demand and adjust services accordingly. This includes optimizing train schedules, ticket pricing, and service levels to meet the evolving needs of passengers. Regular monitoring will ensure the sustainability of the HSR network by aligning with passenger demand and maintaining high levels of ridership.

By following these recommendations, private investors and stakeholders can strategically develop an HSR network that not only meets current transportation needs but also supports long-term economic growth and sustainability across the United States.

Project Status

While Team 54's project does not include the originally planned dashboard mobile app, it is otherwise on track for the Oral Presentation deliverable.

Completed	Next Steps
• Final Report, Executive Summary, and all precursor activities	• Prepare and present Oral Report • Finalize recommendations

References

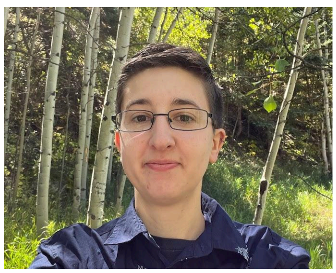
- Bureau of Economic Analysis. (n.d.). *Regional Economic Accounts*. https://apps.bea.gov/regional/downloadzip.htm
- Bureau of Transportation Statistics. (n.d.). *Air Carriers : T-100 Domestic Segment (All Carriers).* https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=GEE
- Bureau of Transportation Statistics. (n.d.). *World Airport Codes List.* https://www.bts.gov/topics/airlines-and-airports/world-airport-codes
- Ewing, R. and Cervero, R. (2001). *Travel and the Built Environment: A Synthesis.* Transportation Research Record 1780. Washington, D.C.: Transportation Research Board, National Research Council. http://depts.washington.edu/trac/concurrency/lit_review/trr1780.pdf
- Federal Railroad Administration. (n.d.). *State Rail Plan Guidance*. https://railroads.dot.gov/elibrary/final-state-rail-plan-guidance
- Hagler, Y., & Todorovich, P. (2009). *Where High-Speed Rail Works Best*. Regional Plan Association. https://rpa.org/work/reports/where-high-speed-rail-works-best
- Klein, B., & Sullivan, K. (2021, November 15). *Biden signs infrastructure bill into law at rare bipartisan gathering* | CNN politics. CNN. https://edition.cnn.com/2021/11/15/politics/biden-signing-ceremony-infrastructure-bill-white-house/index.html
- Rutzen, B. and Walton, C.M. Southwest Region University Transportation Center. Center for Transportation Research. (2011). *High Speed Rail: A Study of International Best Practices and Identification of Opportunities in the U.S*. (Report No. 476660-00071-1). https://rosap.ntl.bts.gov/view/dot/23700.
- Texas A&M Transportation Institute. (2023). *2023 Urban Mobility Report*. https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2023.pdf.
- United States Census Bureau. (n.d). *Metropolitan and Micropolitan - About*. https://www.census.gov/programs-surveys/metro-micro/about.html
- United States Census Bureau. (n.d). *Metropolitan and Micropolitan Statistical Areas Population Totals: 2020-2023*. https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-metro-and-micro-statistical-areas.html

Appendix

Origin	Destination	Distance_km	Duration
Barcelona	Madrid	505	2h45m
Madrid	Sevilla	392	2h35m
Madrid	Valladolid	163	0h55m
Madrid	Valencia	301	1h56m
Beijing	Shanghai	1064	4h52m
Shanghai	Hangzhou	165	1h6m
Shanghai	Wuhan	870	4h30m
Paris	Lille	204	1h10m
Paris	Lyon	394	1h56m
Paris	Bordeaux	499	2h17m
Paris	Marseille	661	3h25m
Tokyo	Osaka	403	2h27m
Tokyo	Hiroshima	816	3h45m
Tokyo	Nagano	178	1h20m
Florence	Rome	231	1h30m

Team 54

Kevin Houseman
Economics,
Business Analytics,
Data Engineer



Tori Sosnowski
Data Analyst,
Visualization

Leslie Stovall
Data Engineer,
Data Scientist

Emma Valind
Systems Analyst,
Analytics Management,
Communications