SOULPAGE
IT SOLUTIONS

A Next-Gen Data Science Company

Use Case

Leveraging Machine Learning to Forecast Commuters' Ridership

Leveraging Machine Learning to Forecast Commuters' Ridership

The demand for public transportation is growing exponentially worldwide. Awareness of global warming and its impact on the Ozone layer depletion due to air pollution caused by an increasing number of personal vehicles, financially feasible transportation, increasing stress levels, and ailments, etc., are a few key factors that individuals consider before opting for public transport. With an increasing number of passengers, it becomes imperative for the transit operators to gather historical data about various aspects and come out with futuristic operations and marketing plans to accommodate new passengers and make profits.

Data Science plays an important role in gathering historical data and extract valuable information that enables organizations to make better decisions. The collection and strategic use of information can enhance forecasting and can help improve the reliability of transport infrastructure by increasing its efficiency and utilization.

Predictive maintenance of the vehicles brings down the vehicles' breakdowns. Real-time data can augment intelligence and manage anomalies by leveraging Data Analytics. Data feeds from the sensors provide details about the traffic congestions, peak and non-peak periods, etc.

Top 5 Best Metro Systems in the World

1. Hong Kong's MTR , Hong Kong.
2. London's Underground, England, United Kingdom.
3. Tokyo's Underground, Japan.
4. Moscow's Metro System, Russia.
5. Seoul Metropolitan, South Korea

Leveraging Data Science in transportation will help in

* Designing and evaluating the crowd reduction strategies
* Analyzing regularity and structure in commuters' travel behavior
* Analyzing the excess journey time and the passenger waiting time

In this scenario, we performed an in-depth analysis of a Transit system in the San Francisco Bay Area and obtained information on the number of commuters, peak and non-peak periods, etc and derived insights from the data analyzed so that the organization can smoothly plan, manage, and evaluate its services.

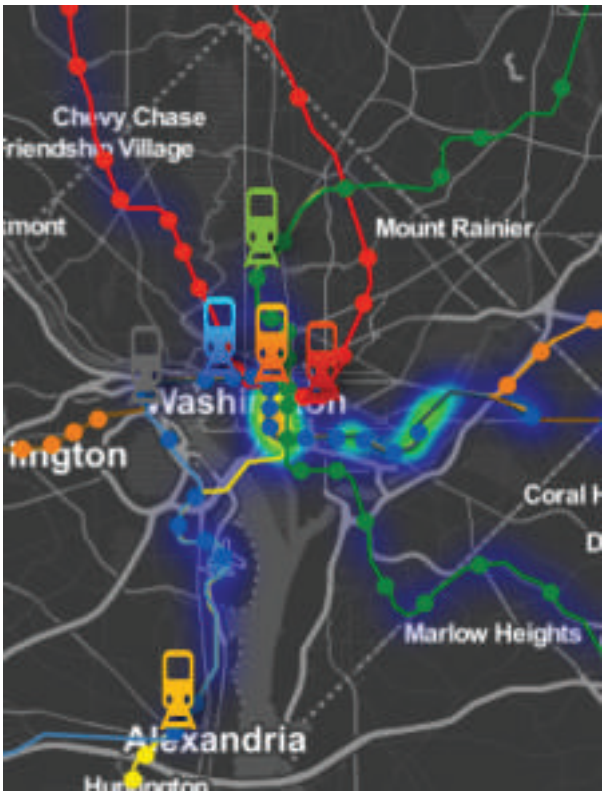
Problem

Bay Area Rapid Transit(BART) is the transit system serving the San Francisco Bay Area and is the fifth-busiest rapid transit system in the United States. BART operates in 6 routes, 46 stations, and covers 112 miles of track.

With a huge passenger count on its side, it becomes difficult for the organization to gain some insights on parameters like the busiest time of the day, the busiest day of the week, the number of passengers who travel during night time, etc. These insights will help the organization to draw relevant marketing and operational plans to meet the demand.

Solution

To meet the growing demands, an in-depth analysis of the weekly average ridership along with the time periods and seasons in San Francisco will provide the organization the required insights to draft their future plans and optimally allocate their resources accordingly.



Technologies Used

We used Machine Learning, specifically Time Series forecasting algorithms, to forecast the average weekly Ridership.

Further, we used the following models to accurately forecast the average weekly ridership.

ARIMA

Holt Winter

About the data set

We gathered 2 data sets of different time periods and features to get a more detailed understanding of the passengers' travel patterns.

Data set 1 contains information on the number of passengers who daily commute through BART for a period covering the entire 2016 to May 2017. This data also includes station-to-station combinations.

Source Kaggle

Data set 2 contains information on average weekly ridership from July 1997 to May 2018.
Source BART.

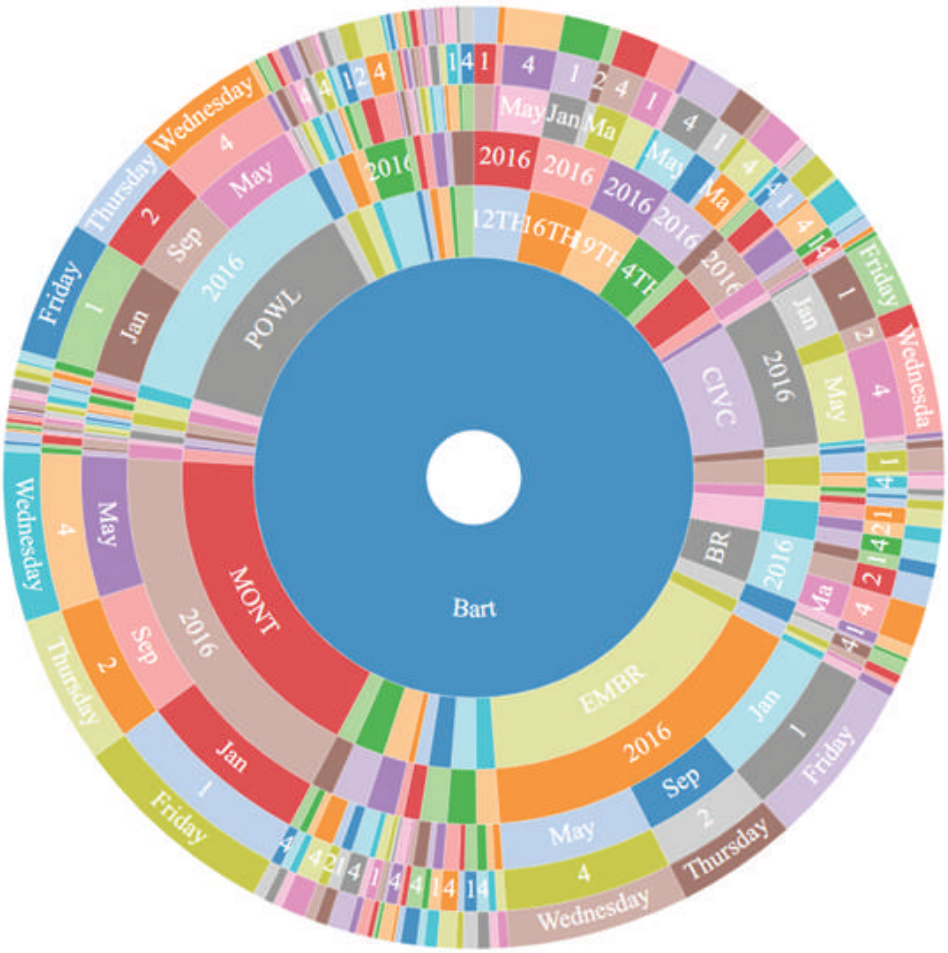
Process

To begin our analysis, we collected all the information from the data sets and cleansed it with respect to the relevant features for our analysis. With this data, we provided station-wise and route-wise insights, that are important for forecasting the weekly rider-ship.

Station-wise Insights

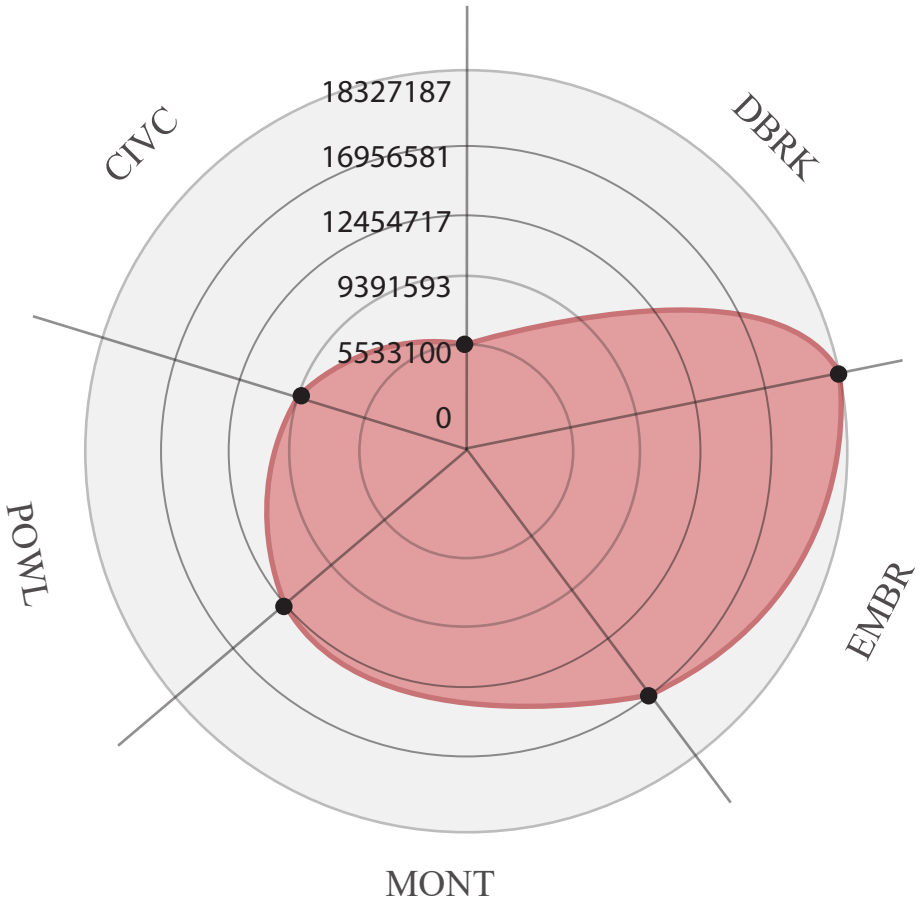
We gathered station-to-station data and analyzed the most populated and least populated stations. The below image gives insights about the overall number of the commuters of a particular station (day-wise,week-wise, month-wise, and year-wise).

Note: The origin and destination stations are given by their respective codes in the plotted graphs.



Commuters count by station & timeline

Top 5 Busiest BART Stations



EMBR:- Embarcadero (EMBR)

Mont:- Montgomery St. (MONT)

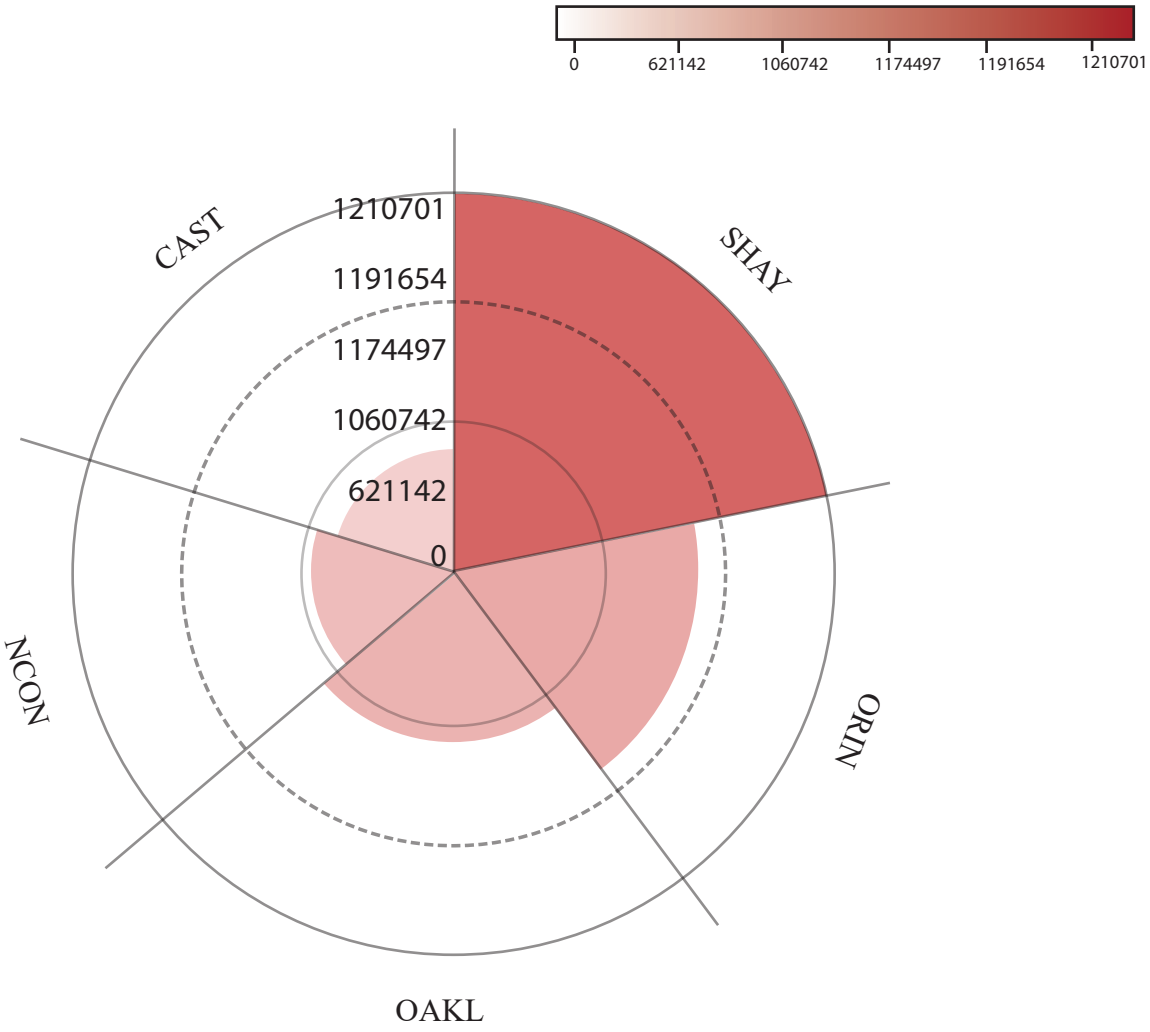
Powl:- Powell St. (POWL)

Civc:- Civic Center/UN Plaza (CIVC)

DBRK:- Downtown Berkeley (DBRK)

Inference:- Embarcadero (EMBR) station was the busiest with the highest number of passengers.

Least 5 Busiest BART Stations



SHAY - South Hayward (SHAY)

ORIN - Orinda (ORIN)

OAKL - Oakland Airport (OAKL)

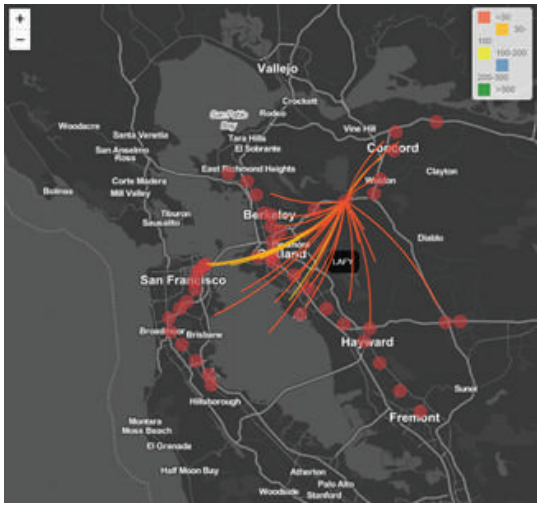
NCON - North Concord/Martinez (NCON)

CAST - Castro Valley (CAST)

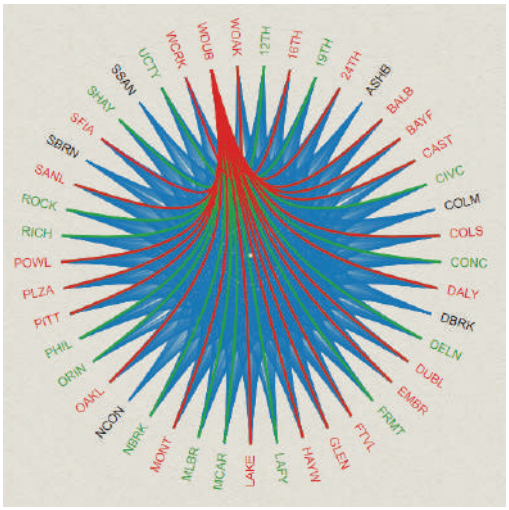
Inference:- The Oakland Airport Station was the least populated station during the period 2016 to May 2017.

Route-wise Insights

The visualizations shown below give information about the routes from one particular origin station to various destinations and the color represents the number of commuters in a particular route.

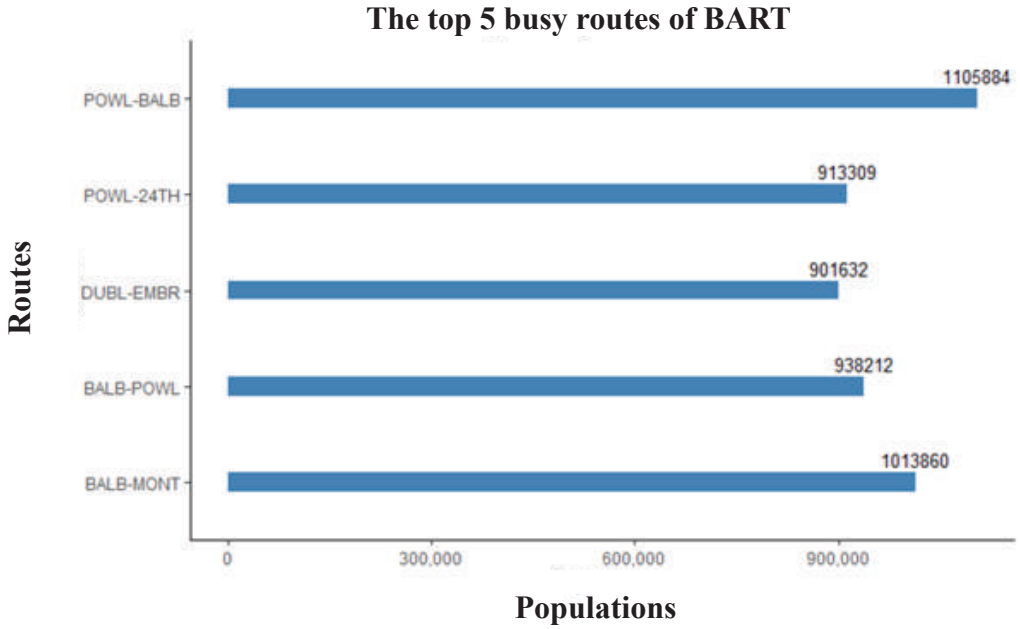


Station to Station Routemap



Commuters Flow chart

From the data sets, we analyzed the busiest and least populated routes.



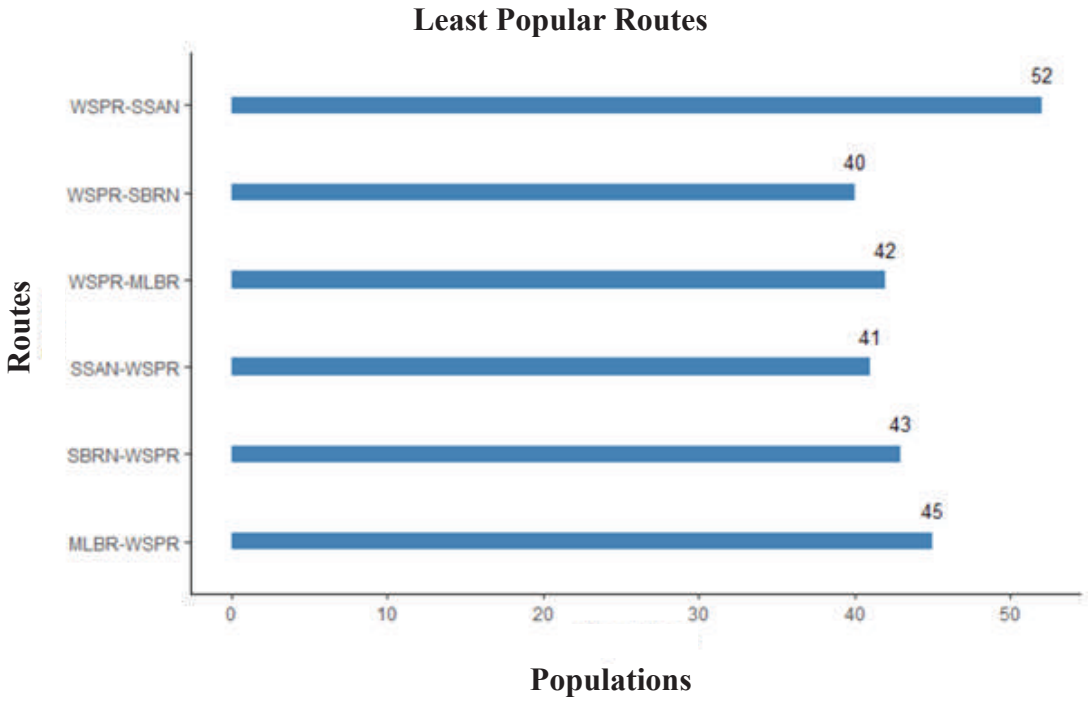
POWL - BALB:- Powell St. (POWL) to Balboa Park (BALB).

POWL - 24th:- Powell St. (POWL) to 24th St. Mission (24TH).

DUBL - EMBR :- Dublin/Pleasanton (DUBL) to Embarcadero (EMBR).

BALB - POWL:- Balboa Park (BALB) to Powell St. (POWL)

Inference:- Powell St. to Balboa Park is the busiest route with over 1105800 passengers in the year 2016 to May 2017.



WSPR-SSAN:- Warm Springs/South Fremont (WARM) to South San Francisco (SSAN)

WSPR-SBRN:- Warm Springs/South Fremont (WARM) to San Bruno (SBRN)

WSPR-MLBR:- Warm Springs/South Fremont (WARM) to Millbrae (MLBR)

SSAN - WSPR:- South San Francisco (SSAN) to Warm Springs/South Fremont (WARM)

SBRN - WSPR:- San Bruno (SBRN) to Warm Springs/South Fremont (WARM)

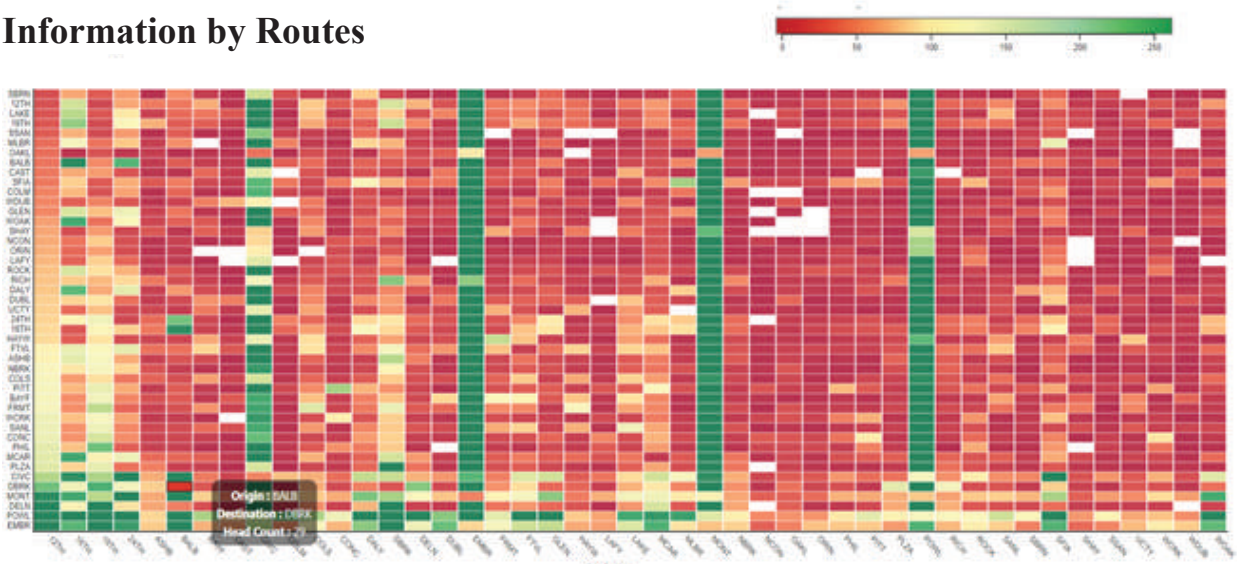
MLBR - WSPR:- Millbrae (MLBR) to Warm Springs/South Fremont (WARM)

Inference:- These are the least popular routes. Only 40 to 50 passengers have traveled on these routes in the time period of 2016 to May of 2017.

Route-wise Insights by Heatmap

The below visualization depicts the number of commuters from the origin station (source) to various destination stations. The shades in the scale represent the magnitude of commuters. The red shade represents the less commuters and the green represents more commuters.

Information by Routes

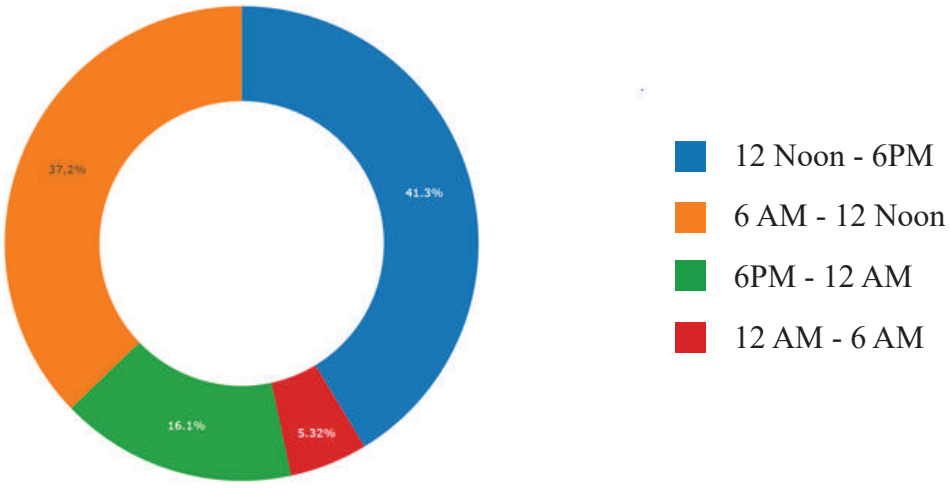


Source

Insights from different Time Intervals

With a 24-hour service rendering organization like BART, it is essential to maintain time intervals for the analysis. We, hence, considered four 6-hour intervals and retrieved information on the percentage of passengers who travel during these intervals.

Percentage of passengers on Time Intervals of the day

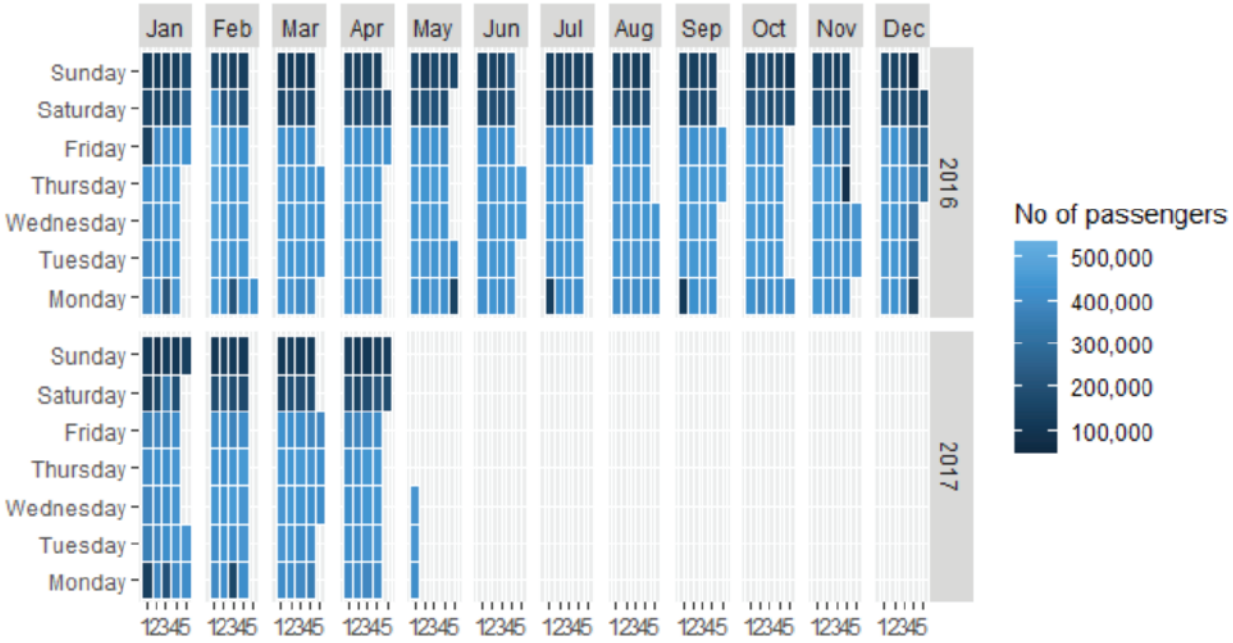


Inference:- It can be seen that 41.3% of passengers have traveled between 12Noon-6pm and 37.2% of passengers have traveled between 6 am to 12 Noon.

Week-wise Insights

The chart below gives information about the number of passengers using BART services in any week of a month. This will provide us with the much-needed numbers about the weekday and weekend commuters.

The Busiest Time



Week of the Month

Inference:- For any public transport system, the number of passengers riding will be more during the weekdays than during the weekends. BART is no different. We can clearly see that from 2016 to May 2017, more passengers travel during the weekdays of any month and during weekends only a few passengers travel for obvious reasons.

Season-wise Insights

Will the seasons affect the travel plans? This was the query that we had since the beginning of our analysis. Hence we considered the following four seasons :

Assumptions

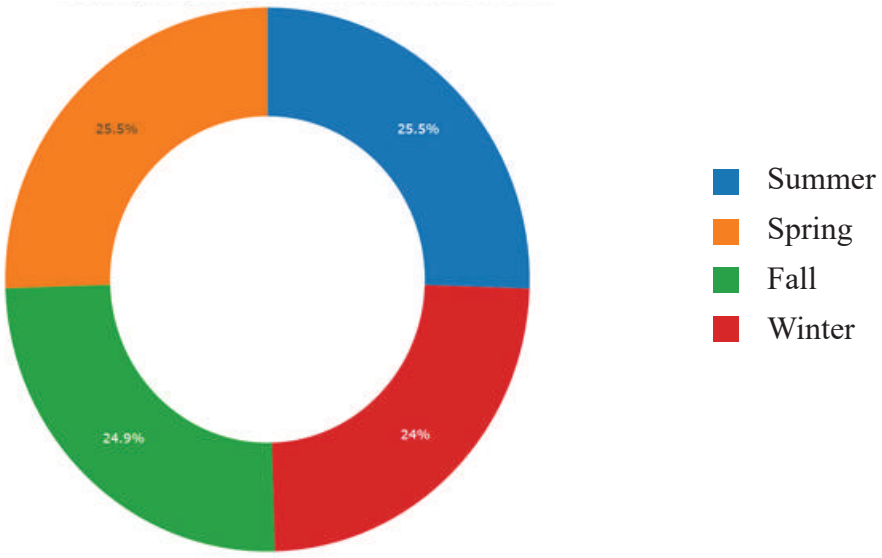
1) Spring:- March, April, May.

2) Summer:- June, July, August.

3) Autum:- September, October, November.

4) Winter:- December, January, February.

Percentage of passengerson different seasons on 2016

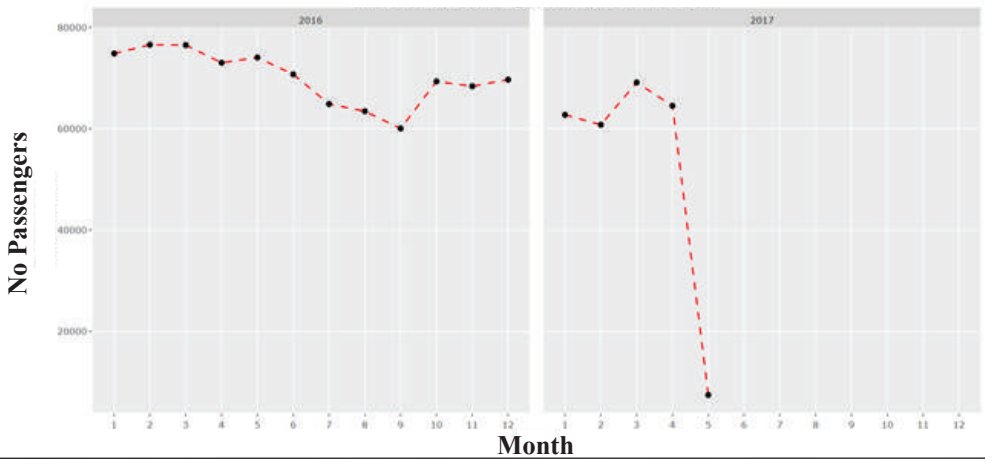


Inference:- We observed that during 2016, seasons did not have any major impact on the number of passengers as the numbers are nearly the same all through.

Insights from the busiest route

POWL - BALB being the most populated route, a thorough analysis of this route will definitely provide valuable insights on risks and opportunities, which can further lead to better decisions.

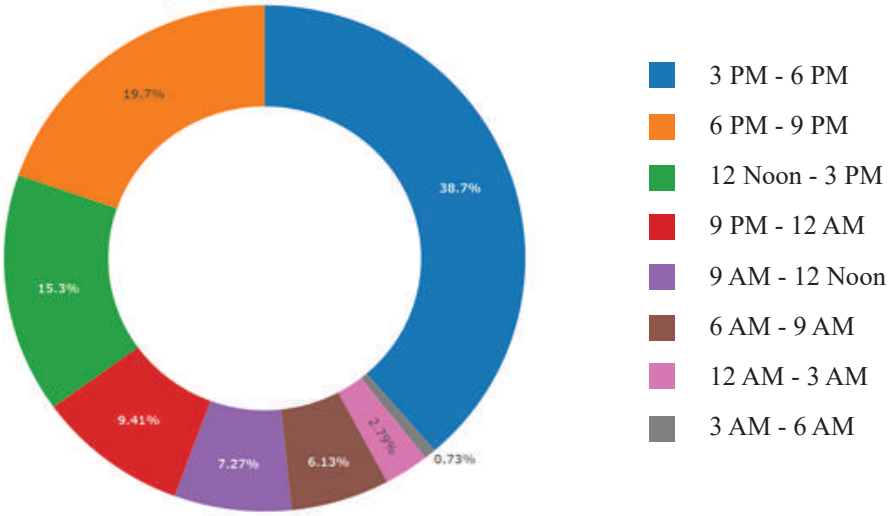
The Trend of Passengers on route POWL- BALB



Inference: - It is observed that during the middle of the year, the number of passengers traveling is falling and starts to increase by the end of the year.

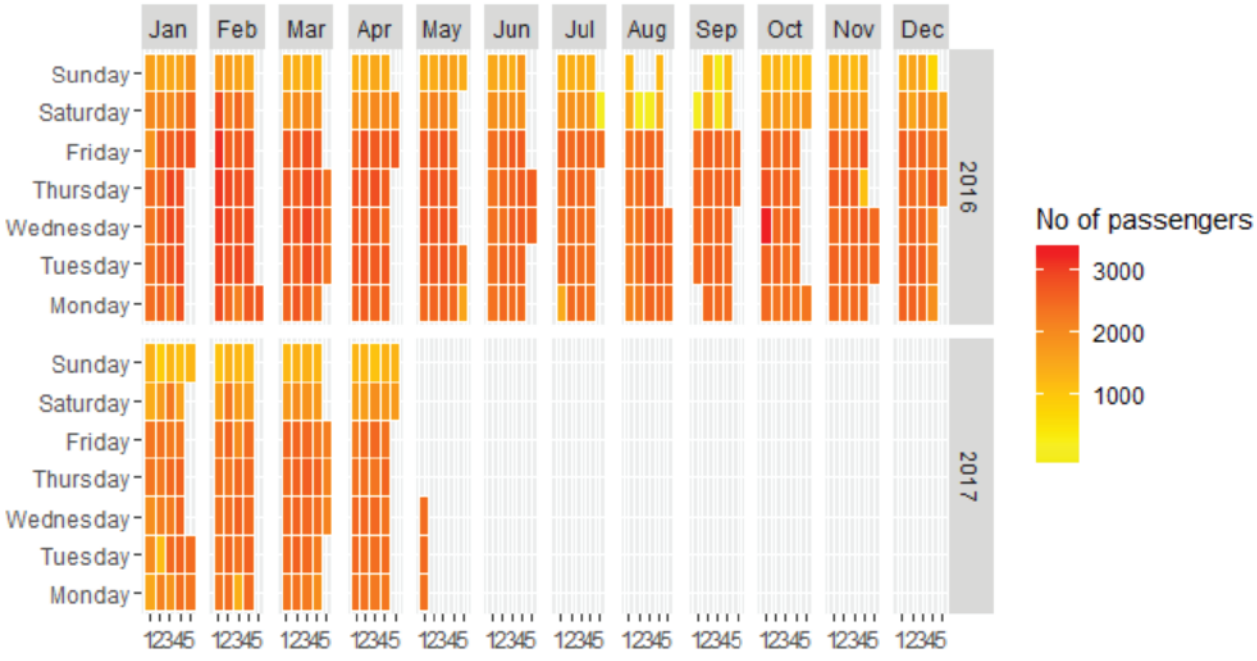
Furthermore, to gain more in-depth knowledge of the travel patterns, we considered the busiest and populated route, which is Powell St. to Balboa Park. We divided the time into shorter intervals (3-hours intervals) to understand the exact time frame when passengers are traveling.

Percentage of passengers at different time intervals



Inference:- 38% of passengers have traveled during the time interval from 3 pm to 6 pm and 19% of people have traveled between 6 pm to 9 pm.

The Busiest time from POWL to BALB

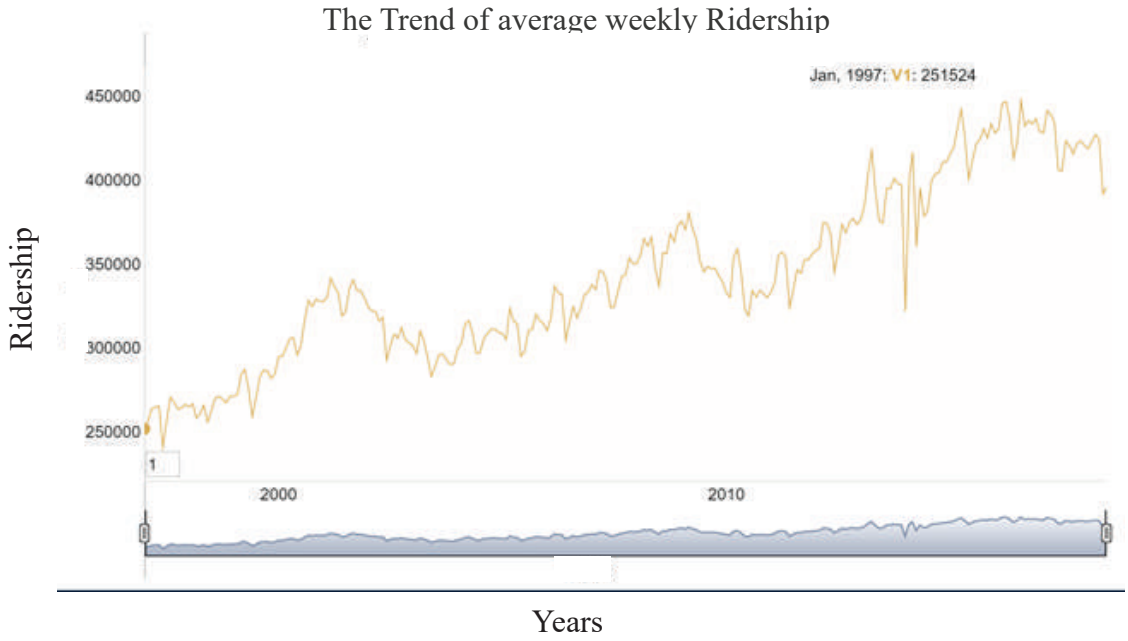


Week of the Month

Inference:- The most populated route of the city commutes with more passengers on weekdays than during weekends. The passenger count during the weekends of August and September is very low.

Forecasting of Average Weekly Ridership

From the data in data set 2 we forecasted the average weekly ridership numbers from August 2018 to July 2020. We used the ARIMA model and the Holt Winter model to predict the average weekly numbers by considering the trend from 1997 to June 2018.



Average Weekly Ridership Predictions

Year 2018

Monthly	Forecast
August	414608
September	404722
October	404229
November	407280
December	407841

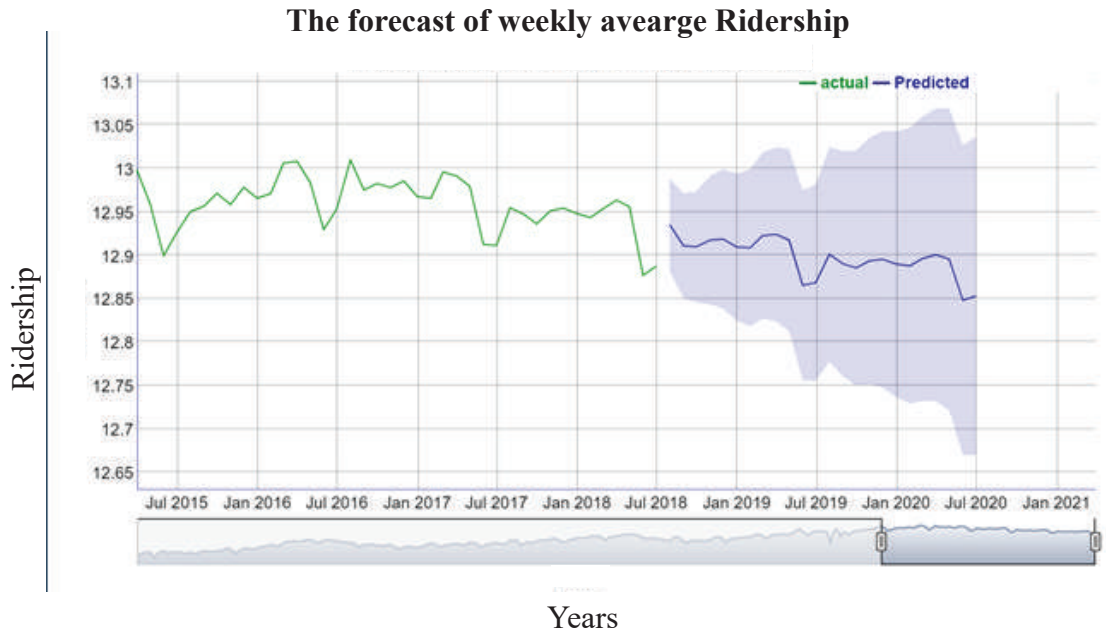
Year 2019

Monthly	Forecast
January 2019	404122
February 2019	403776
March 2019	409430
April, 2019	410061
May 2019	407389
June 2019	386703
July 2019	387868
August 2019	400749
September 2019	396350
October 2019	394525
November 2019	397764
December 2019	398429

Year 2020

Monthly	Forecast
January 2020	396226
February 2020	395458
March 2020	398803
April 2020	400510
May 2020	398489
June 2020	380157
July 2020	381988

Note:- The prediction values are plotted on the graph in log terms.



Conclusion

We have deeply analyzed the data and extracted the insights by which BART organiztion can better plan their resources. Our ARIMA model gave better results for the data that we collected, with an error deviation of 4%. Wehave successfully built a forecasting model and forecasted average weekly ridership from August 2018 to July 2020.

Machine Learning is the most compatible technology when it comes to the transportation industry. The ability to forecast future trends with the help of past data and provide deeper insights that will bring operational and business benefits is the reason for companies to leverage machine learning.

- Demand and supply forecasting - insights about peak and non-peak time of the day, week, month and year, seasonal demand, etc will help in forecasting the demand and drafting the supply strategies to meet those demands.

- On-time performance analysis - can check the performance parameters of the vehicles, which will provide information about the commuters' travel behavior.

- Cost-benefit analysis - the cost and benefit analysis will help the industry to enhance its operational and marketing services in the areas uncovered and provide uninter rupted services for the commuters.

About SoulPage IT

SoulPage IT Solutions is a data science technology company helping businesses derive value from their data. We work with enterprises, startups and entrepreneurs and help them build smart data-driven products & solutions. SoulPage leverages advanced data science technologies like AI, NLP, Machine Learning and Predictive Analytics among others to drive innovation, business transformation and build high-end capabilities for businesses.

To know more about SoulPage, visit https://soulpageit.com/

Our Services

Machine Learning

Text Analytics

Computer Vision

Data Analytics

Digital Transformation

Predictive Analytics

Reinforcement Learning

Recommendation Engine

Our Clients

YITSOL DEERDEN BHANUSAA.COM F CUBE E FACTORS sciens TECHNOLOGIES

Why Choose Us as Your Data Science Partner?

- A dedicated team of expert data scientists with deep authority over AI, Machine learning and NLP, etc.
- Experienced in building AI products and solutions.
- End-to-end services from strategy consulting to software development.
- Complete adherence to data regulations like HIPAA, Data protection 2018, etc.

Contact Us

India

#B13, Indu Fortune Fields Annexe, KPHB 13th phase road, Next to Indu Villas, Hyderabad - 500072
Contact: +91 9032334488, Email: contact@soulpageit.com

United Arab Emirates

H.no 1604, Near KIA Motors, Tourist Club Area, Abu dhabi
Contact : +971 562251149, Email: saikiran@soulpageit.com

Germany

Lütticher Straße 14, 13353 Berlin
Contact : +49 17628069033, Email: dinesh.lanka@soulpageit.com