

BART Density Estimator

Azizi, A., Rejniak, M., Stokely, M.

MIDS W210 Capstone - Final Presentation
April 2023



Team



Ahmad
Azizi



Michael
Rejniak



Murray Stokely

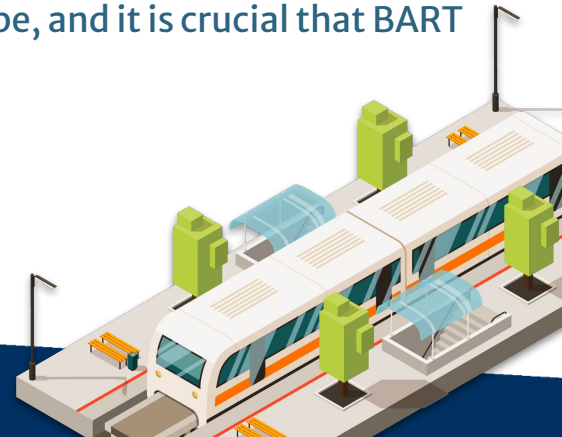


Problem & Solution

Our **mission** is to increase BART ridership through accurate ridership predictions for riders and administrators.

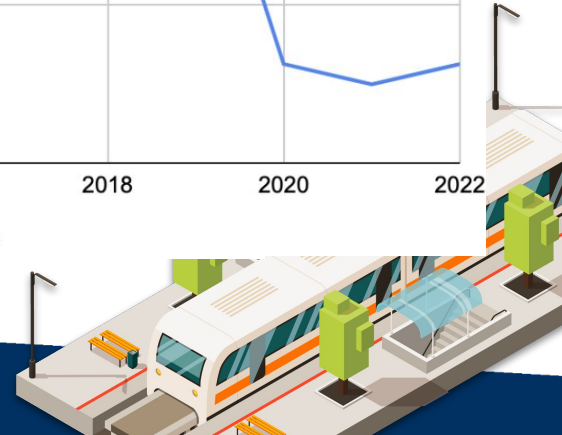
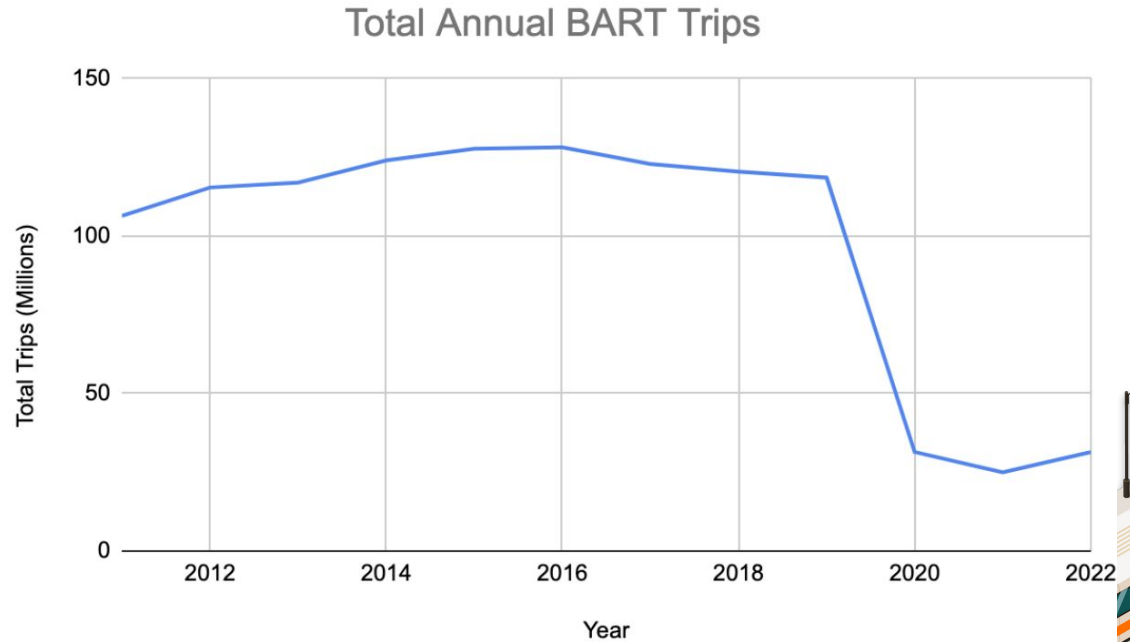
BART ridership was decimated by COVID lockdowns and fails to recover pre-lockdown levels as other mass transit systems are tracking toward.

Transit patterns are changing rapidly in the post-pandemic economic landscape, and it is crucial that BART adapt to meet the needs of the population and encourage more ridership.

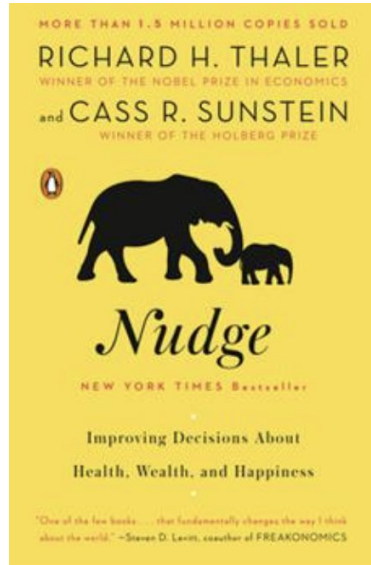
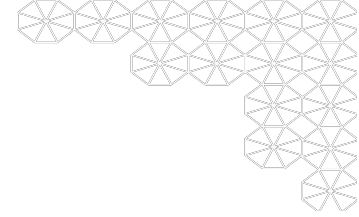


BART Ridership Decline

Significant drop in ridership after COVID-19 Shelter in Place in March 2020.



Encouraging transit riders

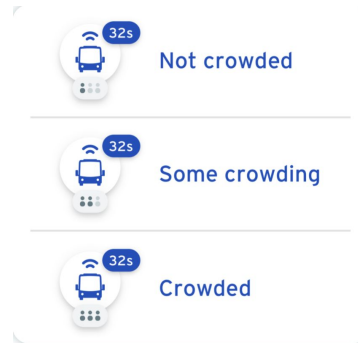
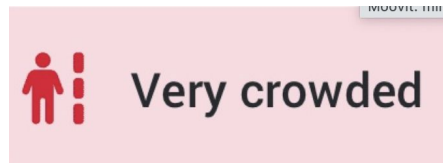
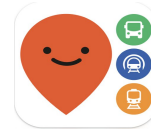


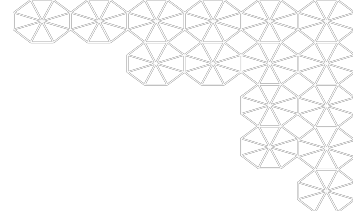
- Can small encouragements or incentives help bring people back to riding transit?
- What if we increased nudges to take transit through mobile apps based on projections of transit density, transit times, to incentive potential riders.



Presenting Density Estimates to Transit Riders

- Existing mobile transit apps rely on user reports.
- Provide real-time, not forward-looking projections of transit density.
- Not useful for planning a trip to avoid crowded trains.





BART DENSITY ESTIMATOR

A Demo



BART Density Estimator

You deserve the most comfortable ride possible when you ride BART. Make sure you have a seat on your next journey! Use the form to predict the number of empty seats on your next journey!



From
Type to search a station...

[Swap stations ↕](#)

To
Type to search a station...

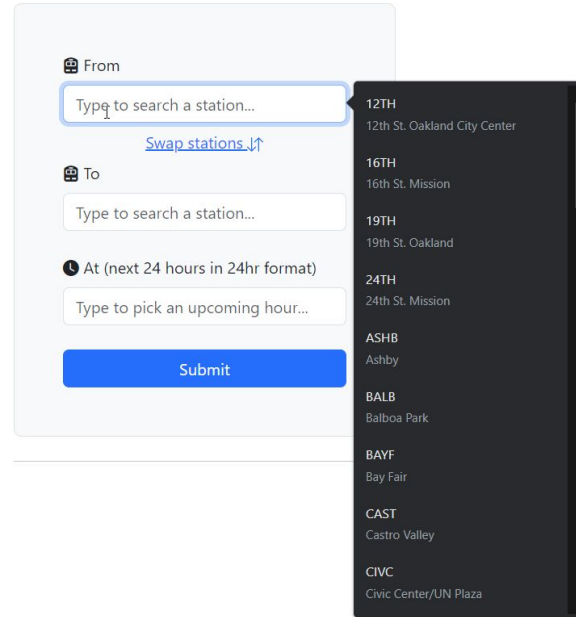
At (next 24 hours in 24hr format)
Type to pick an upcoming hour...

Submit

Live Web Interface - <http://busybart.com>

BART Density Estimator

You deserve the most comfortable ride possible when you ride BART. Make sure you have a seat on your next journey! Use the form to predict the number of empty seats on your next journey!



From

[Swap stations ↕](#)

To

At (next 24 hours in 24hr format)

Submit

- 12TH
12th St. Oakland City Center
- 16TH
16th St. Mission
- 19TH
19th St. Oakland
- 24TH
24th St. Mission
- ASHB
Ashby
- BALB
Balboa Park
- BAYF
Bay Fair
- CAST
Castro Valley
- CIVC
Civic Center/UN Plaza



Signals to Predict BART Ridership



Bart Trips

2011 to present, every hour: number of trips between every combination of stations in the station

1.1 billion trips between 50 stations, represented in 110 million rows in the data set

This dataset is available publicly on the [bart.gov](https://www.bart.gov) website



Stations

Location and id information for BART stations



Weather

Daily weather from 4 locations around the SF Bay Area near existing BART stations going back the last 12 years.

-NOAA



Google Trends

Google searches for “BART schedule” correlate strongly with overall trips, including the significant drop after COVID-19 shelter in place



Events & Holidays

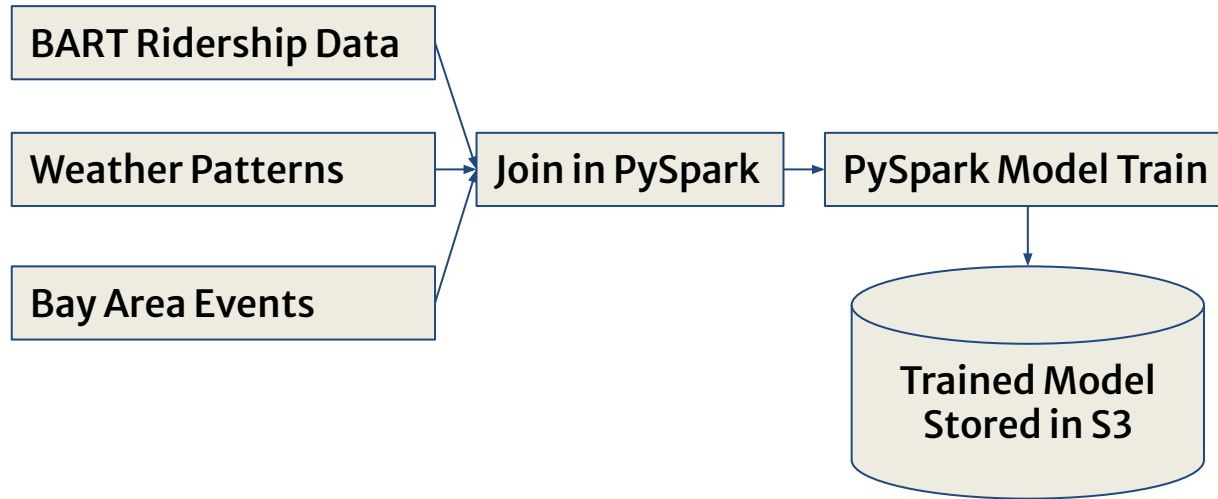
Event data for historic and upcoming events to their closest bart stations.

-dothebay.com



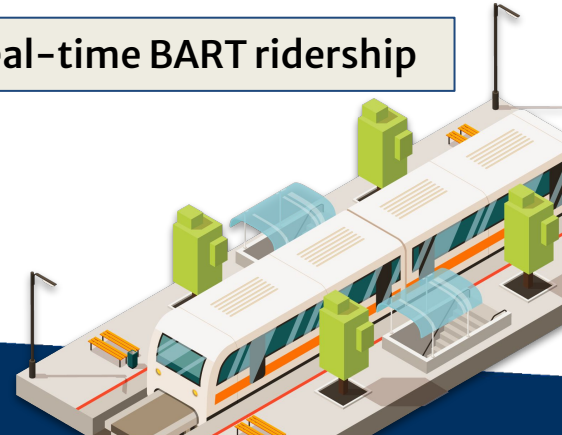
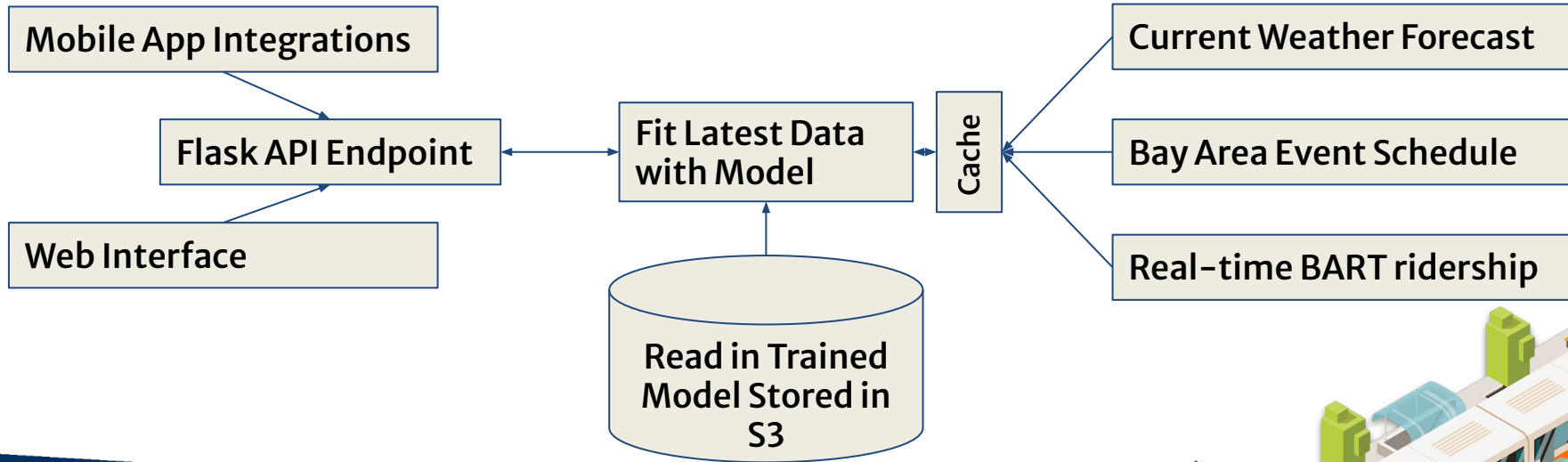
Model Training Pipeline

Weekly offline model training process allows us to update model to reflect changes in ridership and improve predictions for the coming week.



High Level Training and Inference Architecture

API accepts prediction requests, pulls in latest available data on weather and events, and fits the data using pre-trained model in response to user requests.

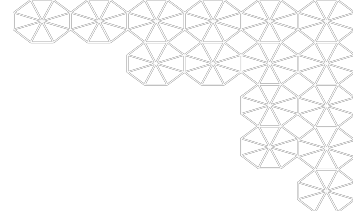


Summary of Key Features of this Design

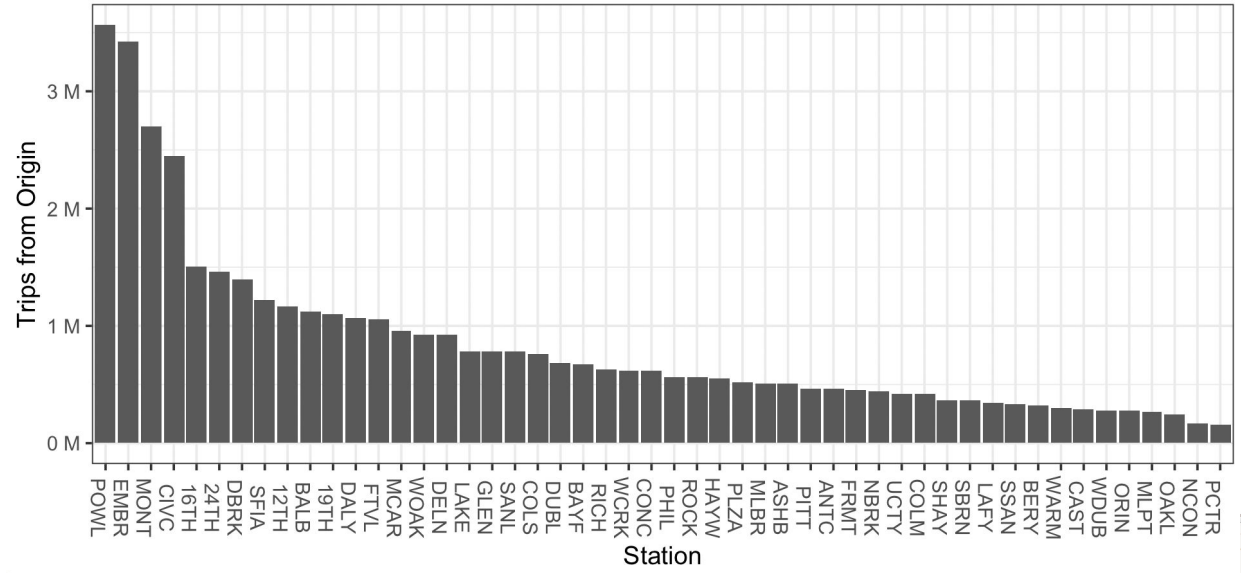
1. Model re-trained weekly over large data set of transit data, weather, events, and tuned to the latest evolution of Bay Area transit trends.
2. Trained model and one-hot encodings of input data sets are serialized and published to the web application server.
3. A simple web API accepts new data from the web or mobile clients over HTTP, encodes the data consistent with how the training data was encoded, then makes a prediction from our fitted model and returns it to the client.



Exploratory Data Analysis



BART Trip destinations led by Trips to/from Downtown SF stations. These stations experienced bigger drop during pandemic.

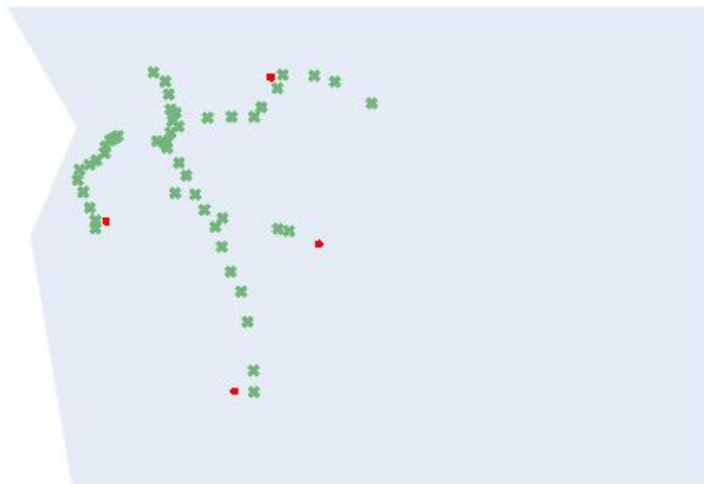


Bart and Weather Stations

- ✦ Bart Stations
- Weather Stations

There are 4 weather stations that encompass the range of all BART stations

- Most BART stations are close to the following stations:
- San Francisco Int. Station
- Concord Buchanan Field



Bart Stations and Venues

We pull event data for historic and upcoming events in the Bay Area

- Sporting events
- Concerts

BART stations near major events may predict increased ridership.



- Bart Stations
- Venues



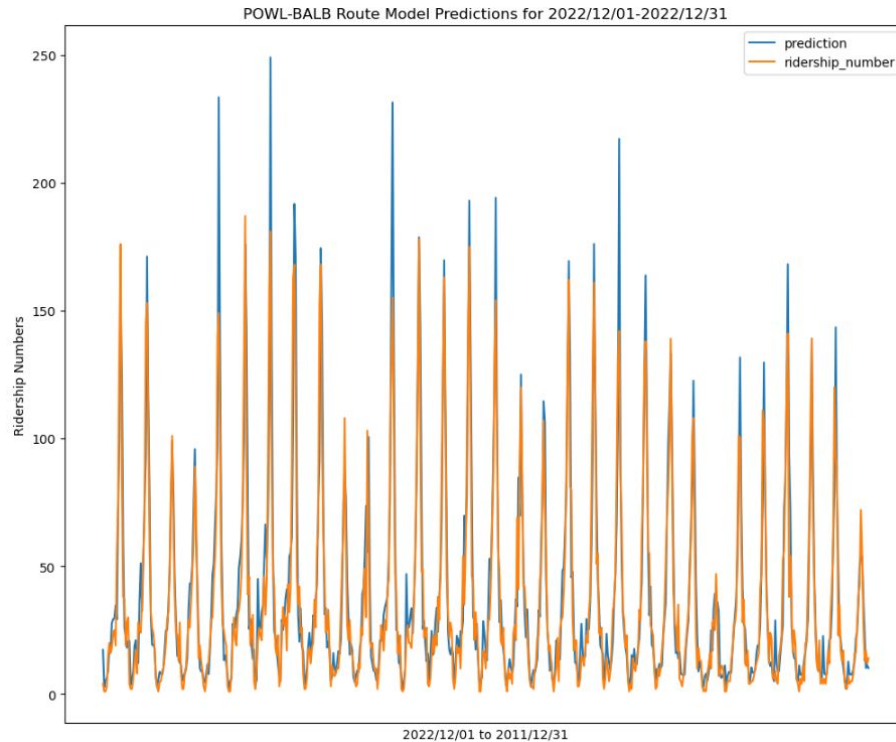
Models

Model Type	RMSE	R2	MAE
Baseline Model (Linear Regression)	0.37	0.28	0.29
Linear Regression	0.33	0.42	0.27
Random Forest Regression	0.27	0.62	0.22
Decision Tree Regression	0.27	0.61	0.22
GBT Regression	0.26	0.73	0.21

Note: The best Model (GBT Regression) was conducted with Cross-Validation

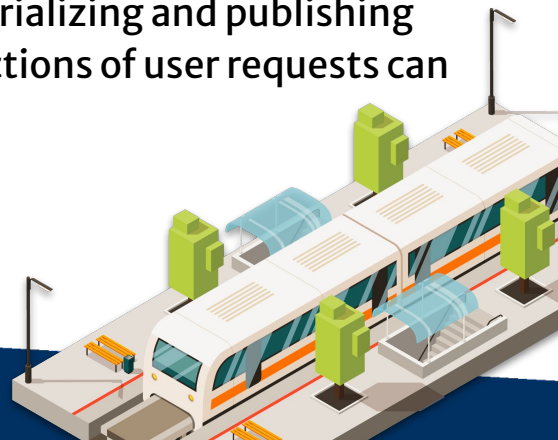


Example Of Model Fit



Summary

- We have created a live web application predicting ridership on BART trips (<http://www.busybart.com>)
- Three technical challenges we overcame:
 - Pulling in high quality weather and event data that can be joined with the BART data to provide high quality predictive signals.
 - Updating our large offline model with the full datasets, and serializing and publishing the trained model to an application server where online predictions of user requests can occur.

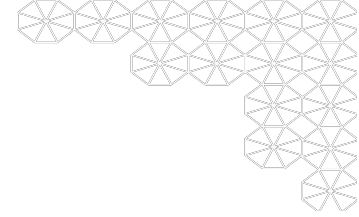


Future Roadmap

- Push our latest development models into our production web application, including the real-time feed of up to the minute BART ridership data.
- Evaluation and Iteration at Scale:
 - We need to invest in validating our predictions at scale with on-site BART trips using the app.
 - Automated comparison of model predictions with data from real-time feed from BART.
- Include visual of results with adjacent hours predictions



Questions



New real-time data sources and the reduction in ridership creates a strong opportunity for innovations in ridership predictions for planning and incentives.

