W261 Section 2 - Group 2

Ken Trinh

Camille Church

Anoop Raj Nair

Shalini Chawla

Dylan Jin

Hector Rincon

Agenda

- ✔ Project Description
- ✔ EDA
- ✔ Modeling Pipeline
- ✔ Next Steps

Project Description

STAKEHOLDERS

Airline Passengers

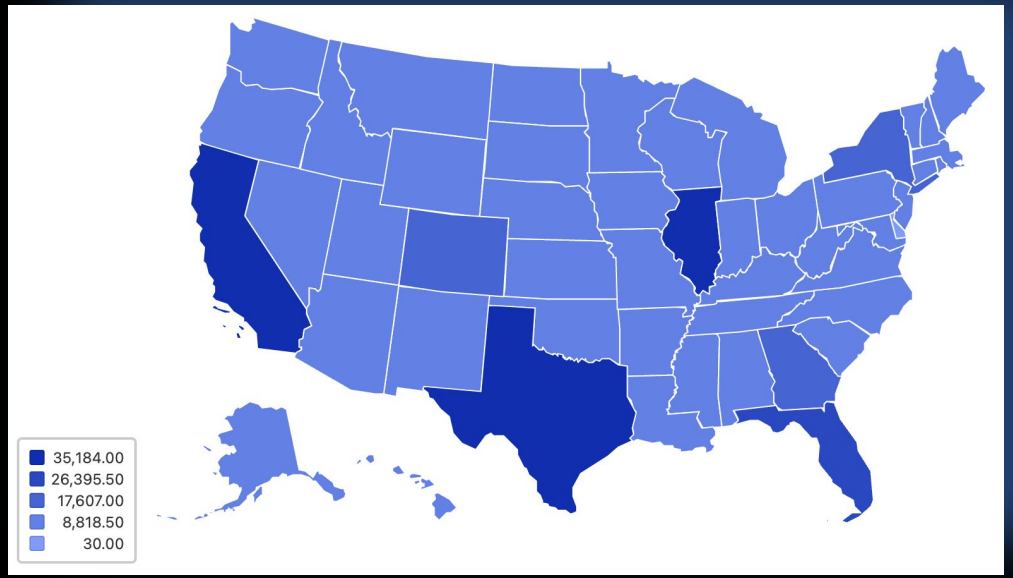
PASSENGERS COST

$18.6
Billion

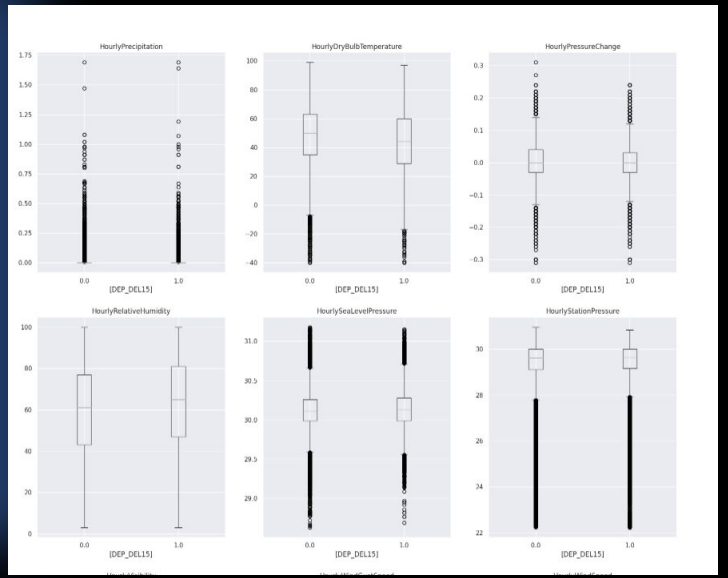
IMPACT TO GDP

$32.9
Billion
Reduced Travel

EDA



Departure Delay Incident Distribution
By State



Departure Delay Distribution
By Weather Factors

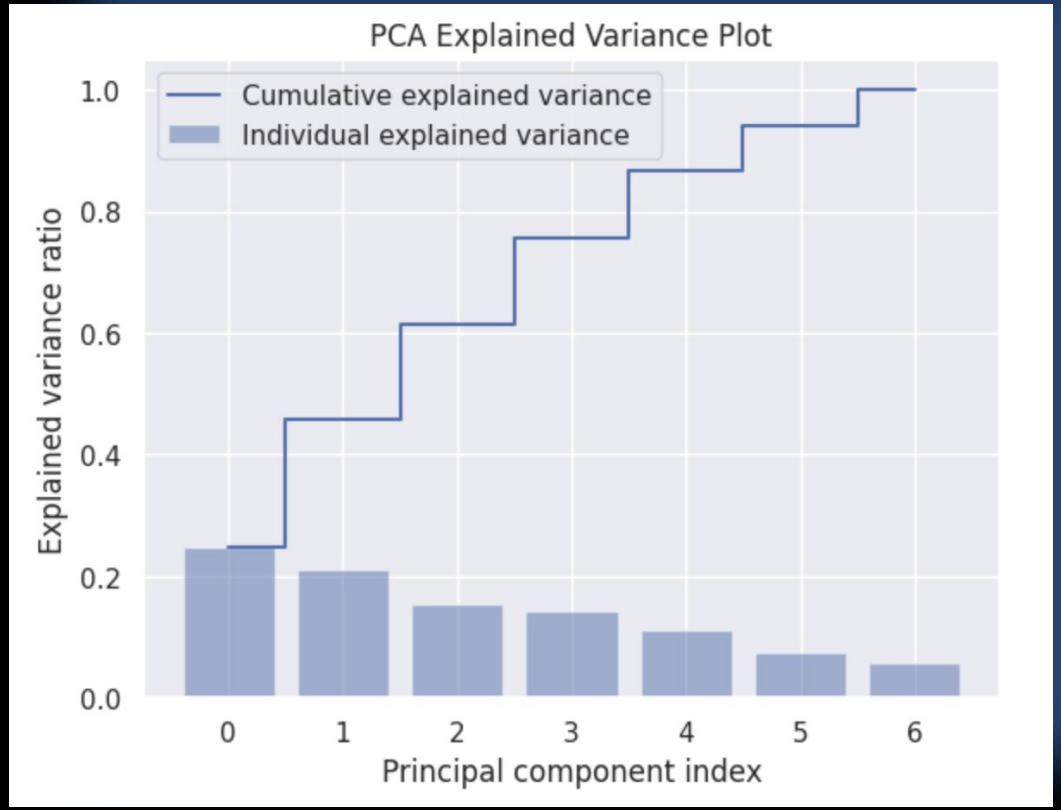
Missing Data

142 columns with missing data

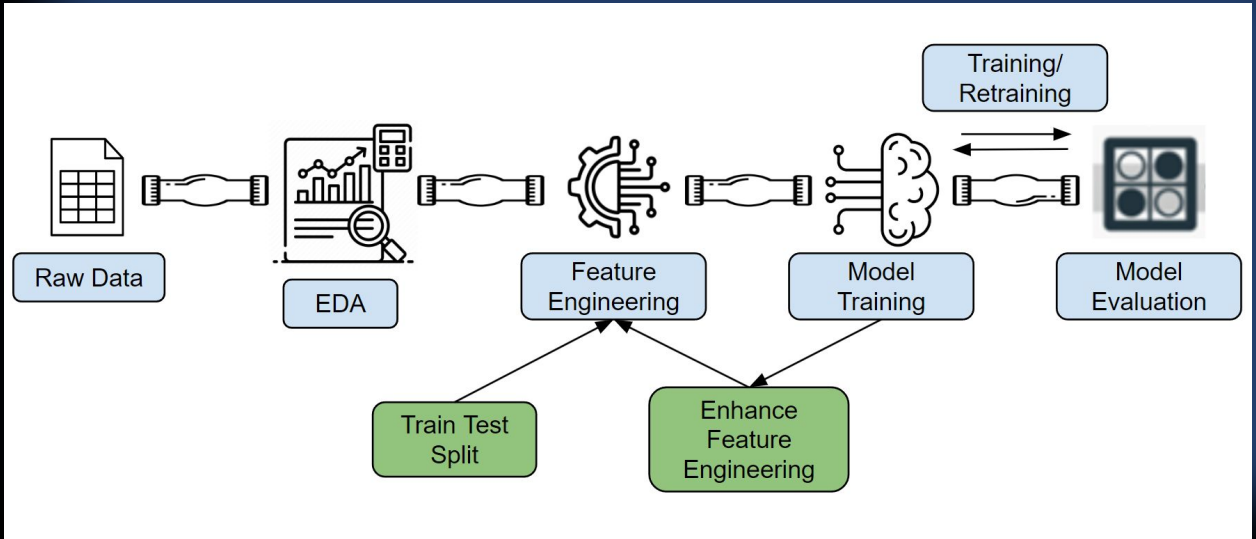
- Discarded 111 columns with > 25% null values
- Ignore flights missing any variables of interest for preliminary analysis and modeling.

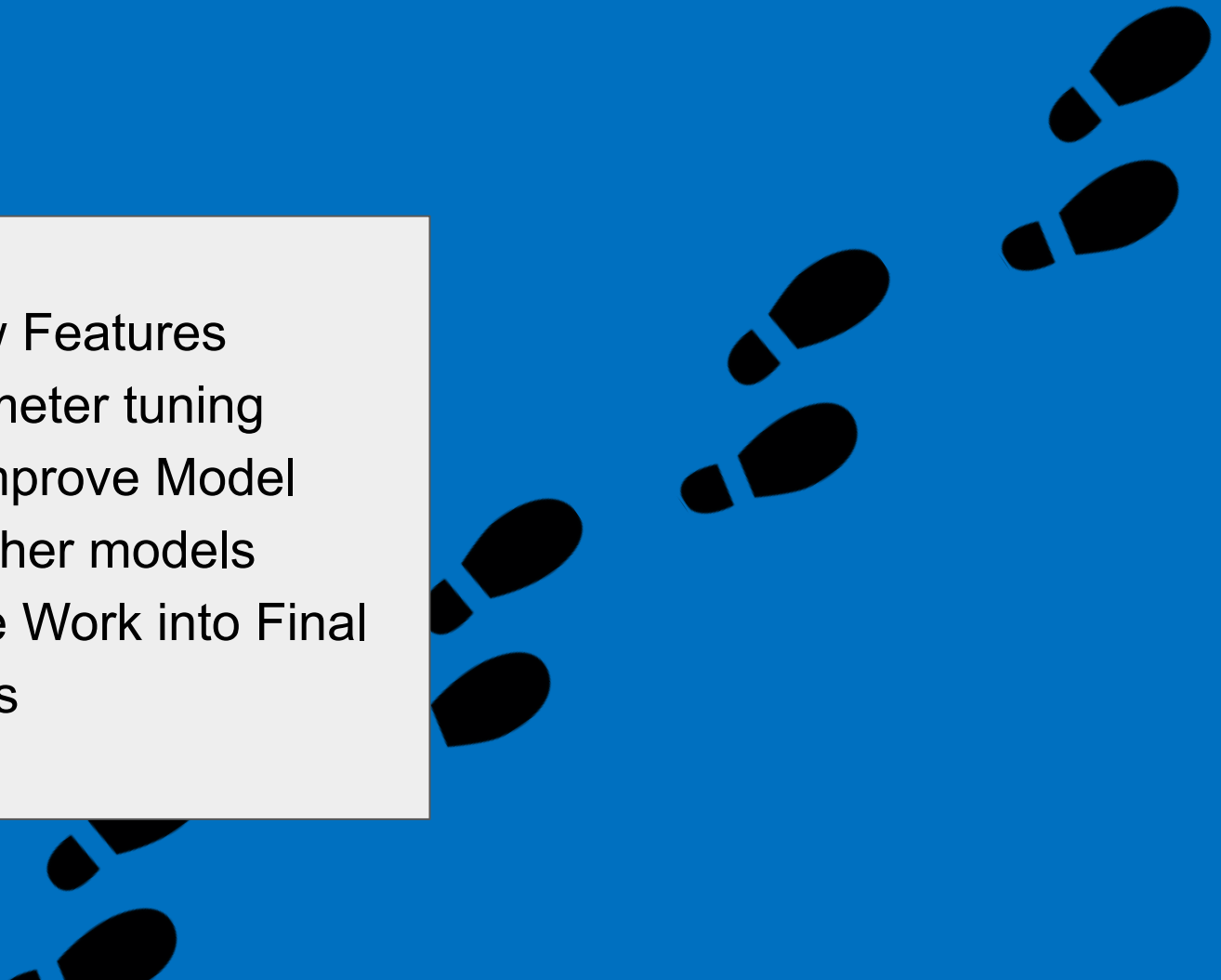
Column	Null Percentage
TAIL_NUM	0.584216937
DEP_TIME	3.018918011
DEP_DELAY	3.018918011
DEP_DELAY_NEW	3.018918011
DEP_DEL15	3.018918011
DEP_DELAY_GROUP	3.018918011
TAXI_OUT	3.076932957
WHEELS_OFF	3.076932957
WHEELS_ON	3.162349798
TAXI_IN	3.162349798
ARR_TIME	3.162349798
ARR_DELAY	3.325405338
ARR_DELAY_NEW	3.325405338
ARR_DEL15	3.325405338
ARR_DELAY_GROUP	3.325405338
CANCELLATION_CODE	96.89951854
CRS_ELAPSED_TIME	0.000142718
ACTUAL_ELAPSED_TIME	3.325405338
AIR_TIME	3.325405338
CARRIER_DELAY	79.58594597
WEATHER_DELAY	79.58594597
NAS_DELAY	79.58594597
SECURITY_DELAY	79.58594597
LATE_AIRCRAFT_DELAY	79.58594597
FIRST_DEP_TIME	99.35940938
TOTAL_ADD_GTIME	99.35940938
LONGEST_ADD_GTIME	99.35940938

Dimensionality Reduction



Machine Learning Pipelines



Next Steps

- Create New Features
- Hyperparameter tuning
- Iterate to Improve Model
- Consider other models
- Consolidate Work into Final Deliverables