

# Blueprint

## Reinforcement Learning for Flight Ticket Pricing

### Raw Data Collection

- Scraping raw data from API i.e Google Flight Pricing Data
- Google flights pricing data in the form of JSON files using the Google QPX Express API service.
- Collected every four hours, date range of 3/2016-3/2017 for one-way flights on the SFO → NYC route.

### Feature Engineering

- Json format (almost 300GB) → Amazon RDS to setup a database and parse appropriate data
- Constructed label (of buy versus wait) for each (flight, date-time) pair
- Normalized features for DQN neural network

- EDA
- Handling Missing data
- Handling imbalanced data
- Cleaning data
- Data Visualization
- Scaling down data
- Converting categorical to numerical data
- Outlier Treatment

### Feature Selection

- For a given (flight, date-time) pair, the information we use includes:
- Flight ID, Source Airport, Destination Airport, Departure Date/Time, Carrier, Current Ticket Sale Price

- Technique to select best feature
- Correlation
  - chiSeq
  - Feature Importance
  - Mutual Information Gain
  - K neighbor

### Model Training

- Training Data: 586 flights, 97,848 data points (65% of the flights)
- Models Constructed:
  - 1) Baseline
  - 2) Q-learning
  - 3) DQN

### Test Model & Evaluation

- Dev Data: 103 flights, 30,451 data points (10% of the flights)
- Test Data: 149 flights, 51,945 data points (25% of flights)

### Hyperparameter Tuning

### Model Deployment