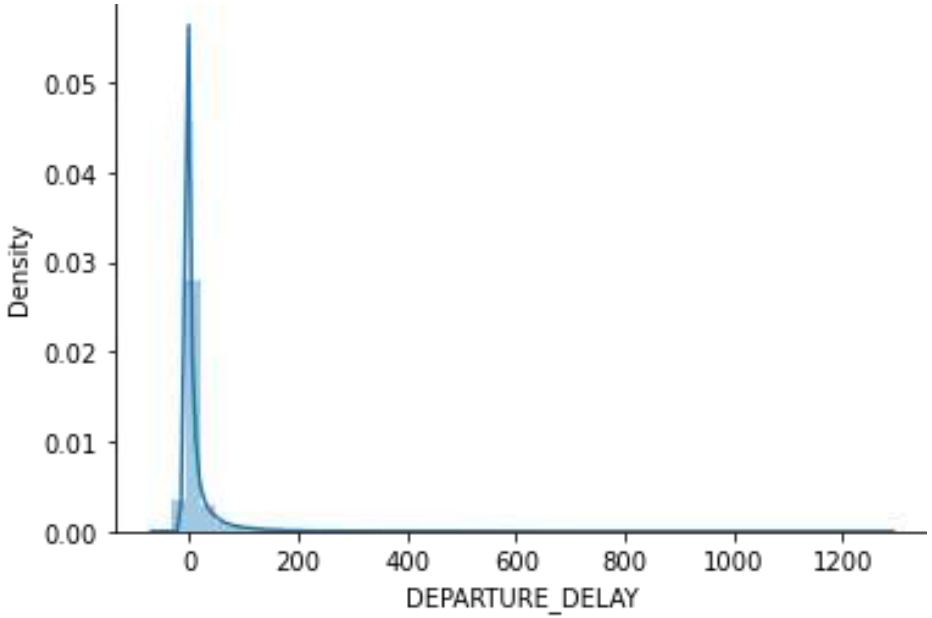
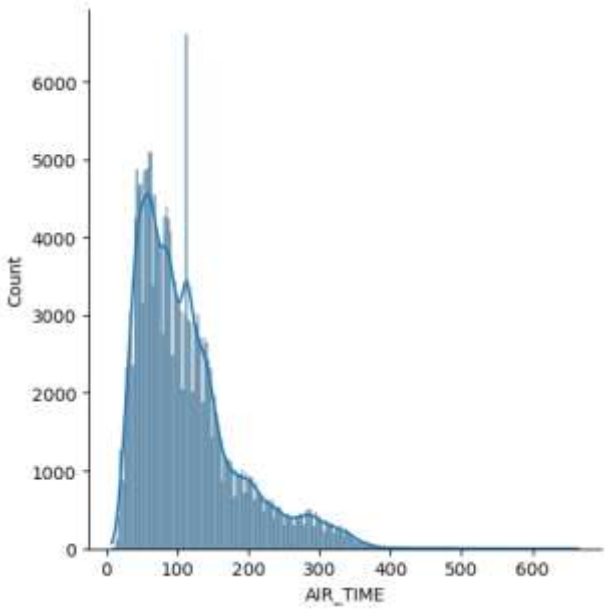
Aviation Industry: AI – ML Case Study

In the aviation industry, delay of a flight can have significant impacts on passengers, airlines, and airports. Predicting the delay time of a flight can help airlines and airports plan accordingly, reduce passenger inconvenience, and improve overall efficiency. The delay of a flight can be influenced by a range of factors including Arrival Delay, Departure Time and the Arrival time. Accurately predicting flight delays is a complex task that involves analyzing a variety of data sources, including origin airport, air time, month, flight number etc. Traditional methods of predicting flight delays often rely on manual analysis of these data sources, which can be time-consuming and prone to errors. Therefore, the challenge is to develop an Auto-ML-based predictive model that can effectively leverage these data sources and capture the complex dynamics of the aviation industry to accurately predict flight delays. We try to develop a predictive model using Auto-ML algorithms that can analyze historical flight data and real-time flight information to predict the likelihood and extent of flight delays. The model should be able to learn from historical data and adjust to changing conditions in real-time to improve the accuracy of its predictions.

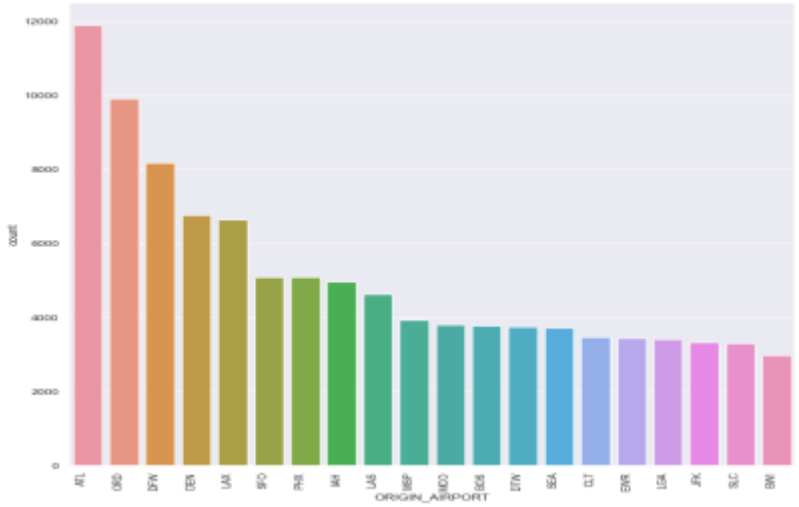


- The feature that we have visualized here is the business feature '**DEPARTURE_DELAY**' in minutes.
- The distribution plot shows that most of the 'Number of trades ranging from -61 to 1284
- Departure delay can be a useful feature in predicting future departure delays for a flight. By analyzing historical data of departure delays for a given flight or route, we can identify patterns and trends in the delays, such as:
    - Seasonal patterns: Delays may be more common during certain times of the year, such as during holidays or inclement weather.
    - Day-of-week patterns: Delays may be more common on certain days of the week, such as Mondays or Fridays when airports are busier.
    - Time-of-day patterns: Delays may be more common during certain times of the day, such as during peak travel times or during rush hour.

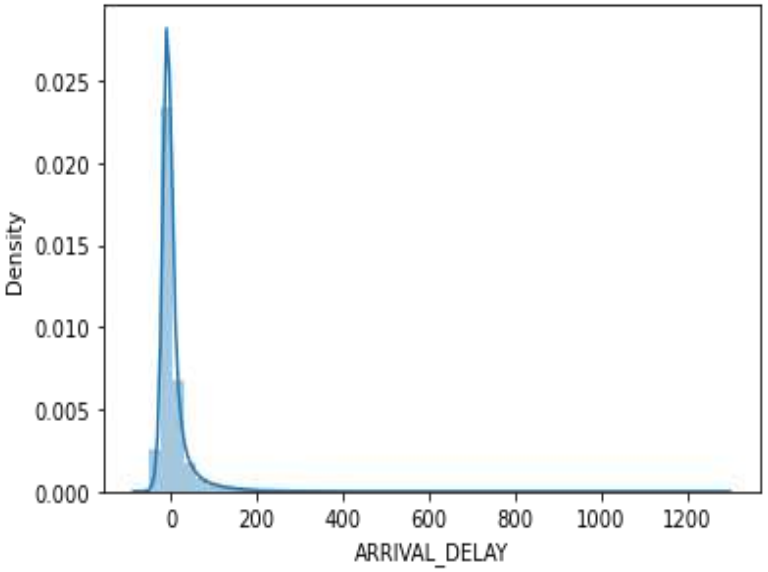
Features Responsible



- The above shows us a distribution plot of the feature "**AIR_TIME**" which the departure time of a flight.
- Air Time value ranges from 0 to 600 minutes.
- Air time is the time a flight spends in the air between takeoff and landing.
- Longer air times can increase the likelihood of delays, diversions, or cancellations due to various factors such as weather conditions, air traffic congestion, and mechanical issues.



- The above shows us a distribution plot of the feature '**ORGIN_AIRPORT**' which is the starting airport.
- Top 20 airports
- The starting point of the flight can affect the flight time, causing delay. Some of the factors involve:
  - The weather conditions of the starting point
  - Flight volume, the number of flights departing from the given airport
  - Airline and airport policies
- Most Departures happen at Atlanta Airport(ATL)



- The above shows us a distribution plot of the feature '**ARRIVAL_DELAY**' which describe 'The time duration elapsed between wheels-on and gate arrival at the destination airport '+ 'The time point that the aircraft's wheels touch on the ground'

Auto-ML Methodology Results

Algorithms	25 percentiles	50 percentiles	75 percentiles	90 percentiles
Lasso	88.2	91.2	92	95.2
Random Forest	92.4	94.5	95.5	96.5
XGBoost	93.8	94.9	97.1	98.5
MLP	93.4	93.2	97.2	97.2
RNN	90.3	90.1	64.9	70.5
Total Features	6	11	17	20
Avg. Accuracy in percentile	**91.62**	**92.78**	**89.34**	**91.58**

- Based on our observation , XGBoost was the best performing algorithm with 98.5% accuracy in 90th percentile
- 50^{th} percentile is the best percentile with an average accuracy of 92.78%.

Conclusion

Flight departure delay prediction is a crucial aspect of the aviation industry as it helps airlines and other stakeholders to plan and manage their operations more efficiently. The dataset contains features of the flight including information about origin airport, destination airport, airplane information, departure time and arrival time, delay time etc. The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset. The dataset has 5819079 records with 5 Categorical Features and 26 Numerical Features. Out of 5819079 records a random sample of 200000 records is selected for modelling.

For regression, models were created with algorithms using Auto-ML techniques like Lasso, Recurrent Neural Network, Multilayer Perceptron, Random forest and XGBoost . With these models, performance measurement values were obtained for feature sets of 6, 11, 17 and 20. The Auto-ML algorithms were able to predict flights departure delays with an average accuracy between 88% – 93% and helped to identify factors that determine the flight departure delays. The major factors include Air Time, Origin Airports and Arrival Delay. The Random forest with 96.5 % accuracy in 90^{th} percentile where tree showed a threshold of arrival delay <= 76.5 mins which can reduce the departure delay. By accurately predicting departure delays, airlines and airports can take proactive measures to minimize the impact of delays on passenger satisfaction, safety, and resource allocation. Based on the performance measurement values obtained, it is possible to say that the study achieved success in predicting in flight departure delay.