

Data-Driven Optimization for Airline Operations



by

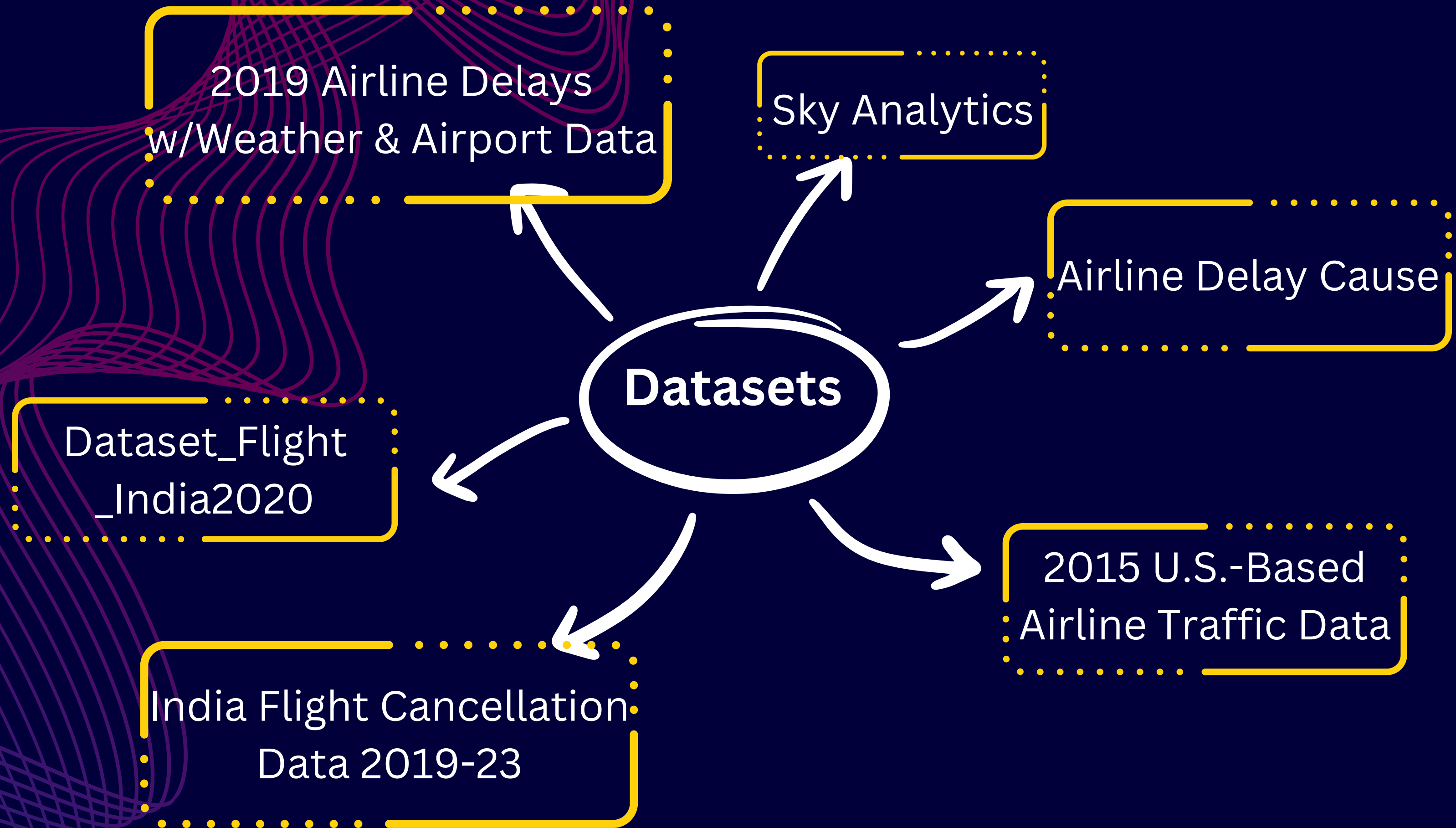
D-42

APPROACH

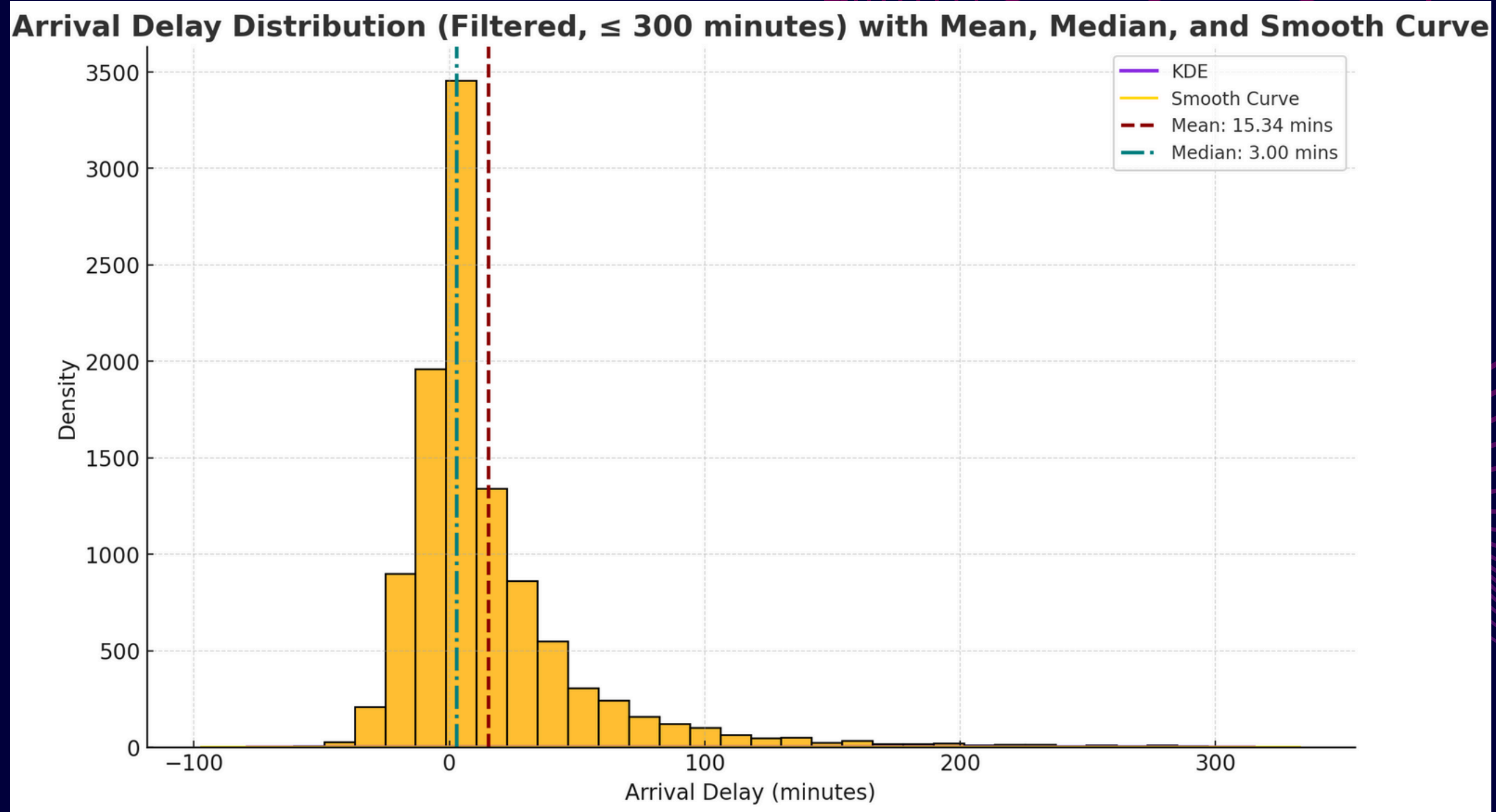
Collective Data from various Sources
and we choose the best one



DATA CURATION

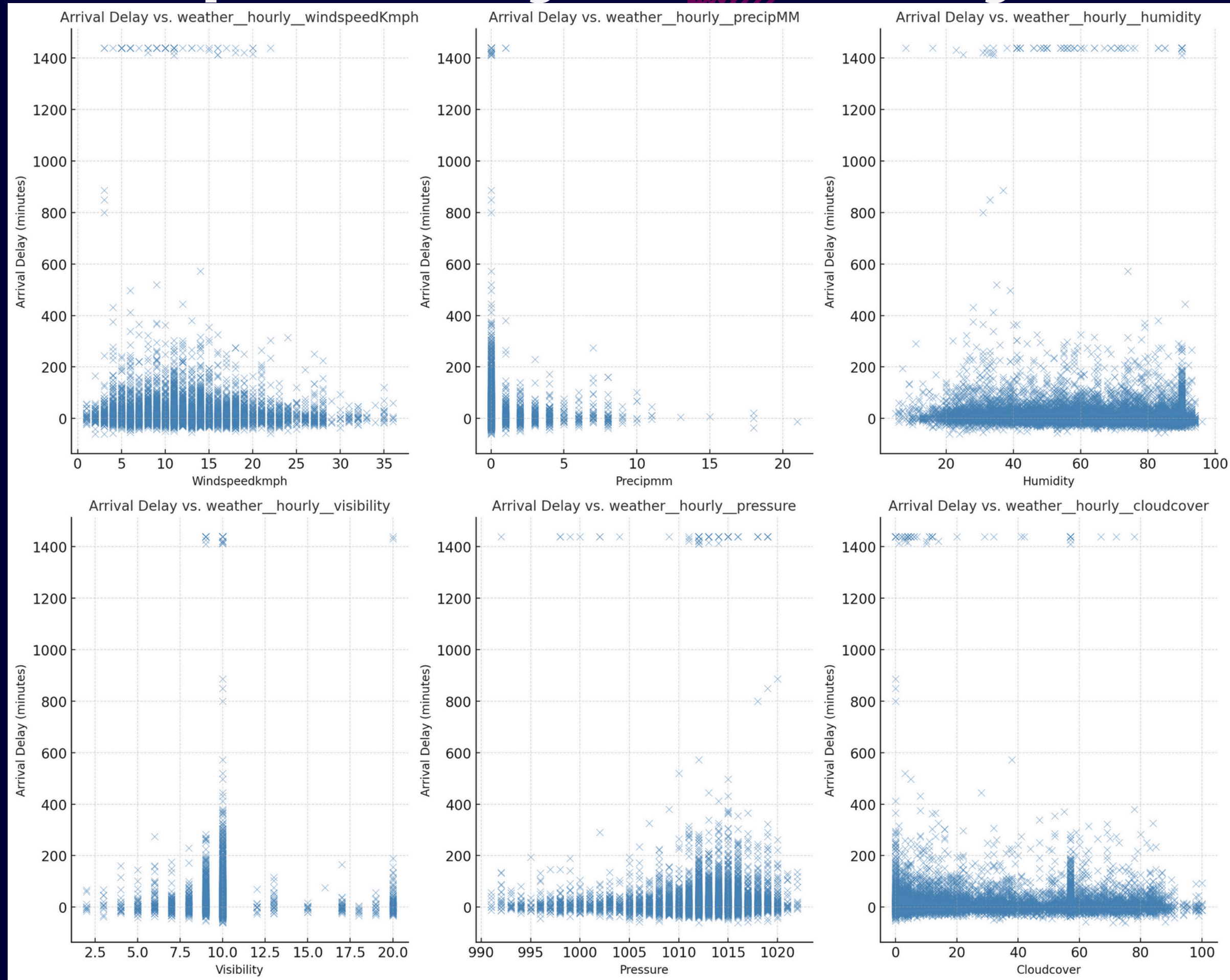


Exploratory Data Analysis

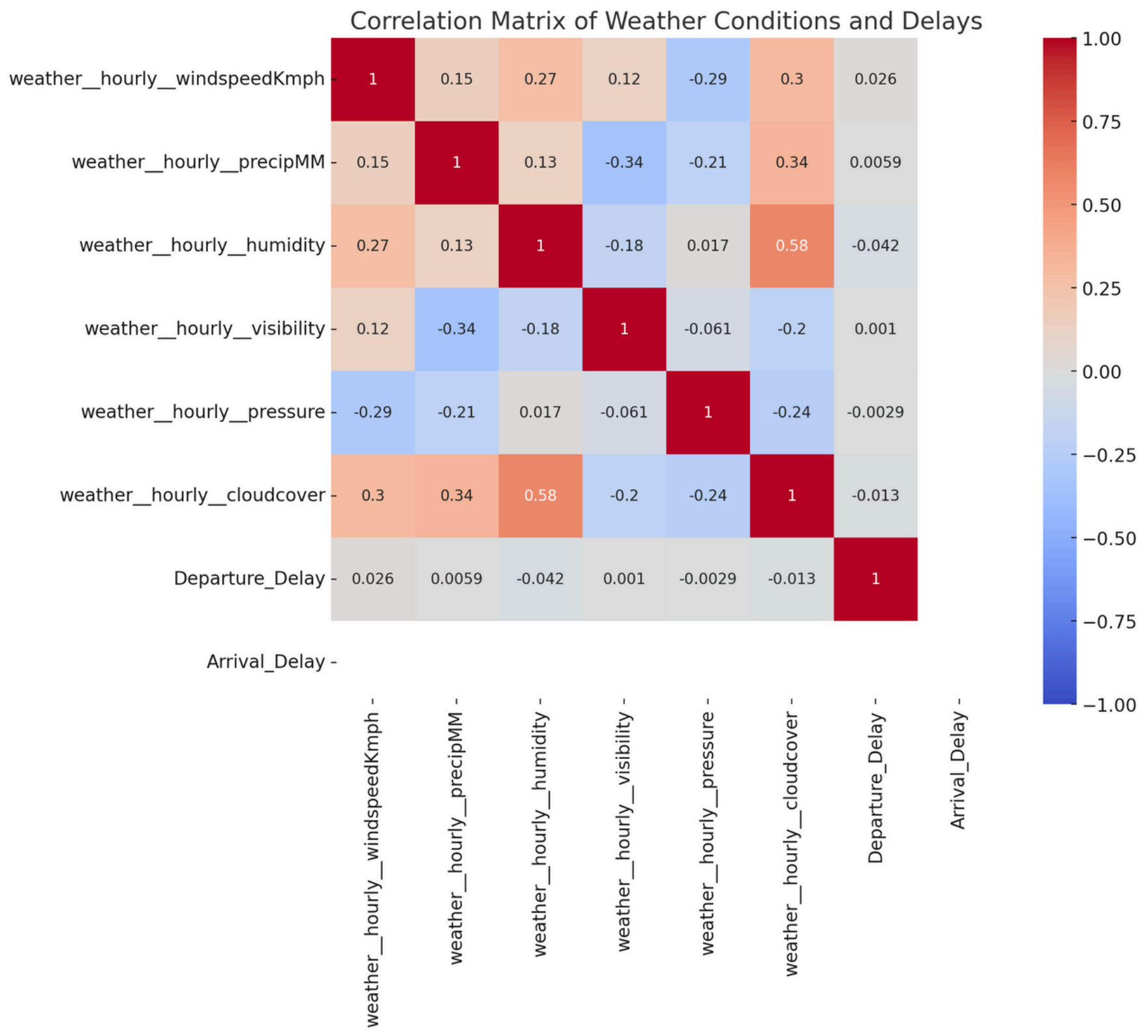


Frequency Distribution of Arival Delay

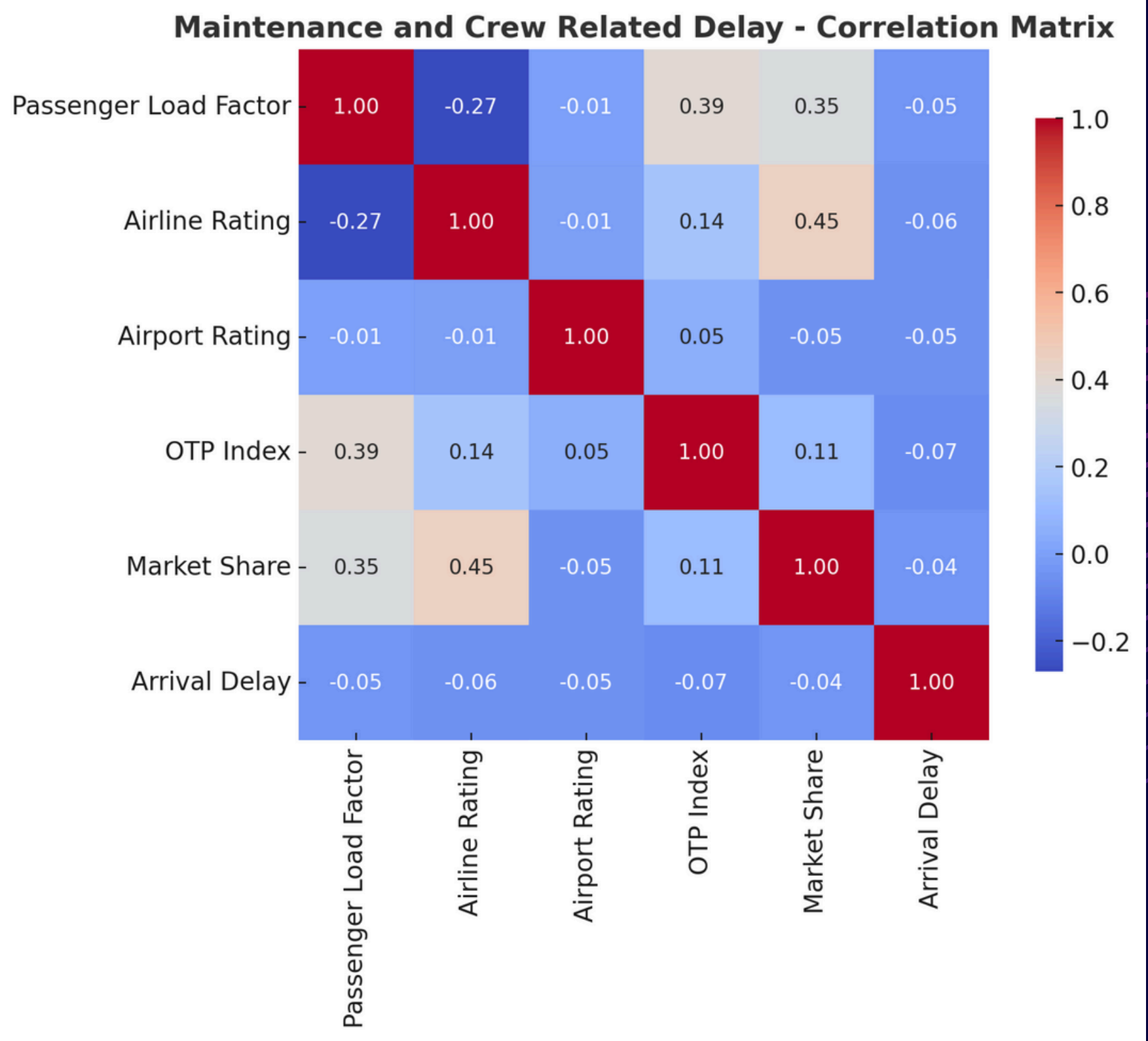
Exploratory Data Analysis



Weather Delay



Maintenance & Crew Related Delay



Feature Engineering

1. Remove columns with $|\text{correlation}| > 0.7$ based on the correlation matrix.
2. Perform PCA, but low accuracy due to weak feature correlations.
3. Apply one-hot encoding for categorical features.



Weather Related Features

- WindspeedKmph
- weatherhourlyprecip
- Humidity
- Hourlyvisibility
- Pressure
- Cloudcover
- Category



Maintenance & Crew Related Delay

- Passenger Load Factor
- Airline Rating
- Airport Rating
- Market Share
- Distance
- OTP Index

Model Selection & Hyperparameter Tuning

Classical Model:

- Linear Regression
- Foundation for linear relationships but underperformed with complex data.

Neural Networks:

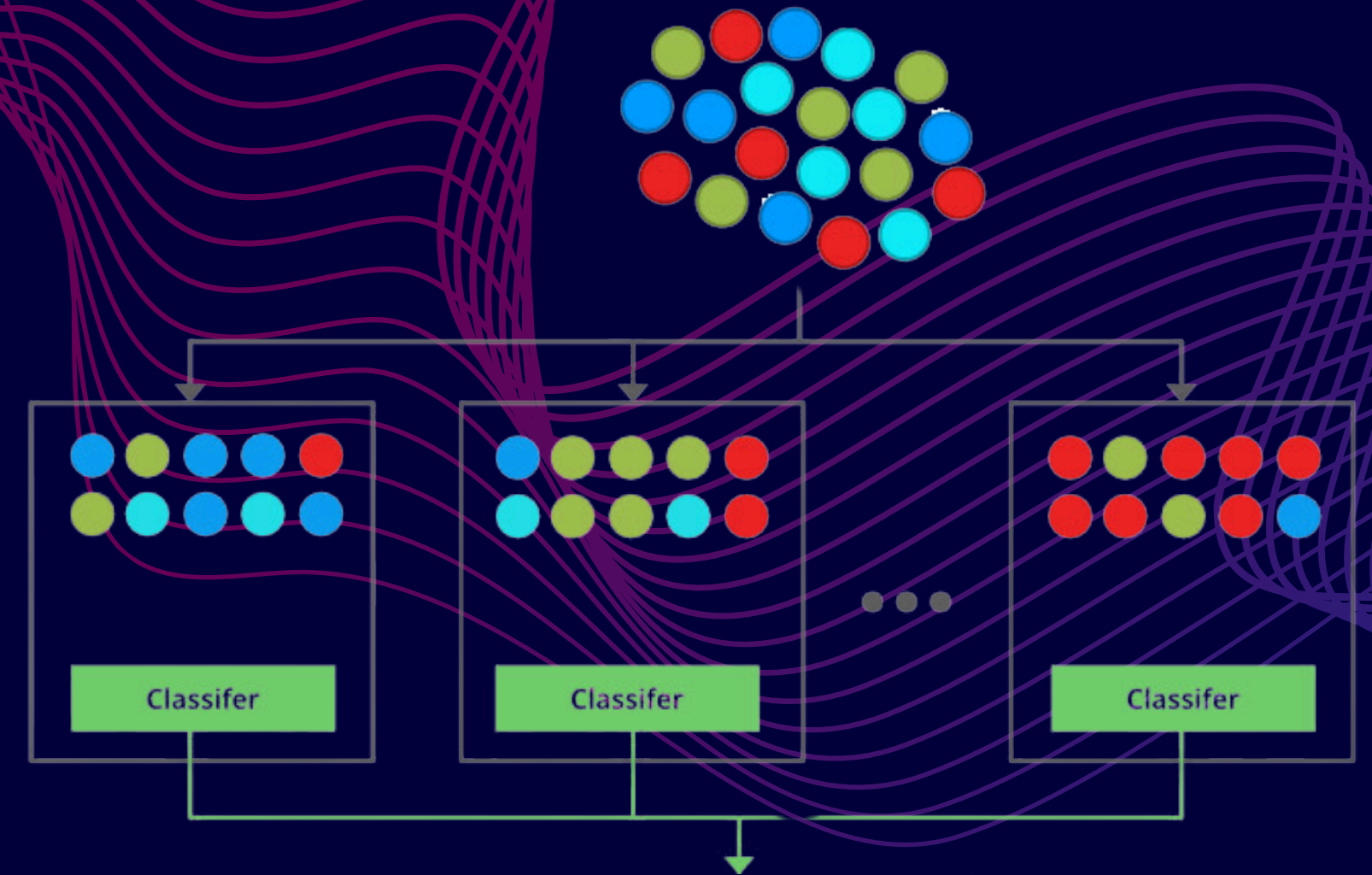
- MLP Regressor
- Designed for nonlinear patterns, but our dataset is too small for effective neural network performance.

Ensemble Methods:

- XGBoost: High accuracy, captured intricate patterns.
- Random Forest: Balanced accuracy with interpretability, handled mixed data types effectively.

Best Model: Bagging

- Bagging (Bootstrap Aggregating) emerged as the top-performing model, reducing variance and improving accuracy, ideal for stable delay prediction.



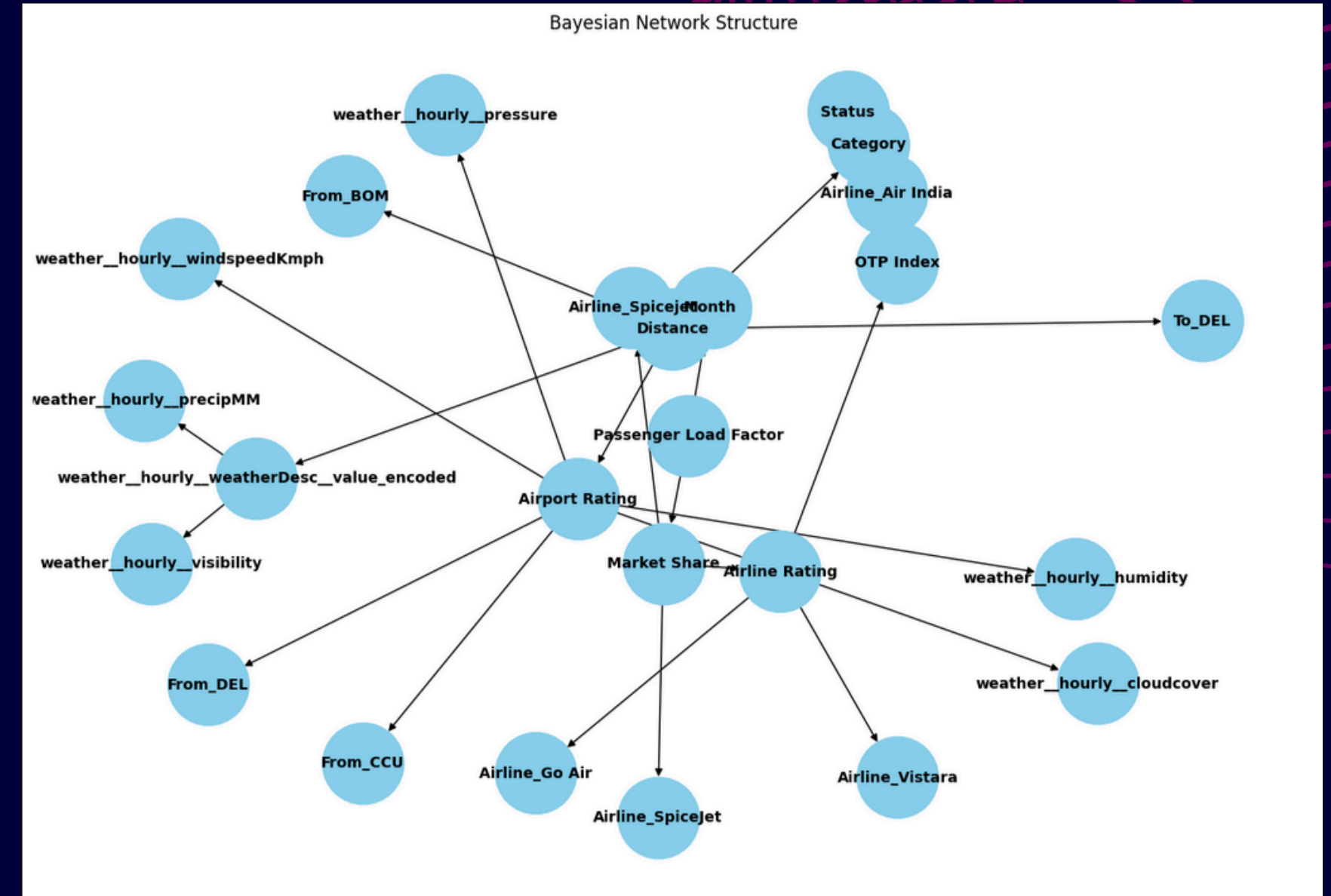
RESULTS

Model	Adjusted R-Squared	R-Squared	RMSE
BaggingRegressor	0.9661524556232209	0.9665789062523265	19.513566008334678
RandomForestRegressor	0.9626515074630269	0.9631220670983504	20.49791073006251
ExtraTreesRegressor	0.962263529914206	0.9627389777407653	20.60410219713155
XGBRegressor	0.9621232544091338	0.9626004695892334	20.64236767625923
LGBMRegressor	0.9454766265781511	0.9461635752866858	24.76648753155644
HistGradientBoostingRegressor	0.9454678460195383	0.9461549053557363	24.768481674868422
DecisionTreeRegressor	0.9398953313405086	0.9406525996810622	26.003221916590032
GradientBoostingRegressor	0.9395383590758823	0.9403001249671334	26.08032647021062
KNeighborsRegressor	0.936713108094016	0.937510469774586	26.682710386729635
MLPRegressor	0.9346856567785087	0.9355085626427085	27.106743155408015
AdaBoostRegressor	0.9319129515175149	0.9327707911390861	27.676126872107357
ExtraTreeRegressor	0.9206164275433387	0.921616593878537	29.883975958146237
BayesianRidge	0.9163044453610839	0.917358939049955	30.684867657968343
Ridge	0.9163024848604077	0.9173570032499405	30.68522703996763
RidgeCV	0.9163024848603701	0.9173570032499035	30.685227039974492
TransformedTargetRegressor	0.9162862749492462	0.9173409975700443	30.68819833918945
LinearRegression	0.9162862749492462	0.9173409975700443	30.68819833918945
LassoLarsCV	0.9162603699050725	0.9173154189076684	30.692946172602937
LarsCV	0.9162272814939827	0.9172827473827659	30.699009493616195
LassoCV	0.9162119898323474	0.9172676483832232	30.7018112258119
LassoLarsIC	0.9161914219731457	0.9172473396617715	30.705579257666766
OrthogonalMatchingPursuitCV	0.9159389601810976	0.9169980586762494	30.751792644035202
SGDRegressor	0.9156022329388878	0.9166655739144595	30.81332303449081
OrthogonalMatchingPursuit	0.9155589865862891	0.916622872429579	30.82121656780148
LassoLars	0.9154188261735784	0.9164844779203415	30.846785442041522

Delay Categorization

Bayesian Network Model for Delay Prediction

- Structure Learning:
- Visualization:
- Model Fitting:
- Inference:
- Output:
 - Display the most likely delay reason and its probability for each row to reveal probable delay causes.



Rescheduling



Initial Strategy:

- Delay Classification
- Consistent Adjustments
- Weather & Load Optimization



Actual Strategy – Predictive Modeling & Iterative Optimization:

- Data Preparation & Classification
- Predictive Modeling & Rescheduling
- Sequential Application

THANK YOU!