Business Analytics:
Prescriptive and Predictive Analytics

Optimizing Delivery Routes at DHL

Presented by: Pharaoh Kipkirui Chirchir

21 Jan 2024

Statement of the problem

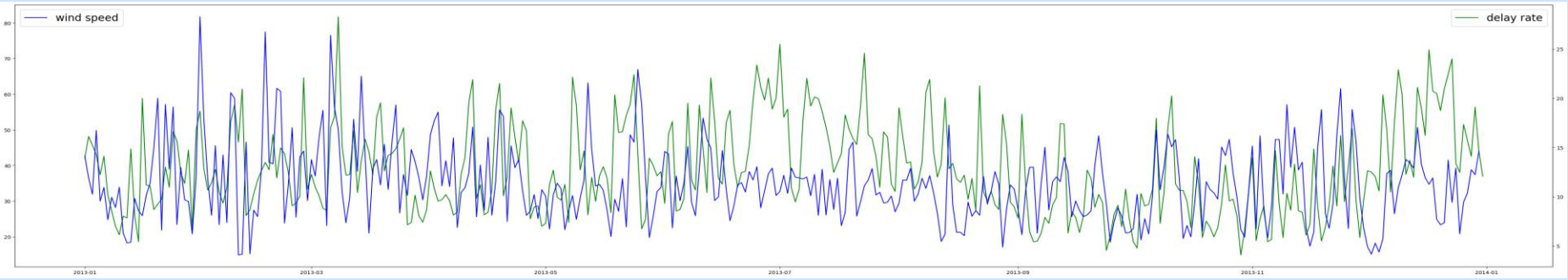
- Logistic industry facilitates seamless movement of goods and services across different regions by means of air, water and sea.
- DHL is a multinational logistic company operating in over 220 countries.
- Route optimization reduces transportation and maintenance cost
- Efficient route planning ensure timely deliveries, which is crucial in ensuring that customer satisfaction is met.
- This presentation focuses at predictive and prescriptive analytics tools that DHL can employ to achieve its goals.
- For real life scenario, Kaggle Dataset on **Flights Delay Classification on NYC Flights** data is used to analyze the causes of flight delays, predict future trends and the current action that needs to be addressed.

Key Factors Contributing to Fuel Consumption and Travel Time

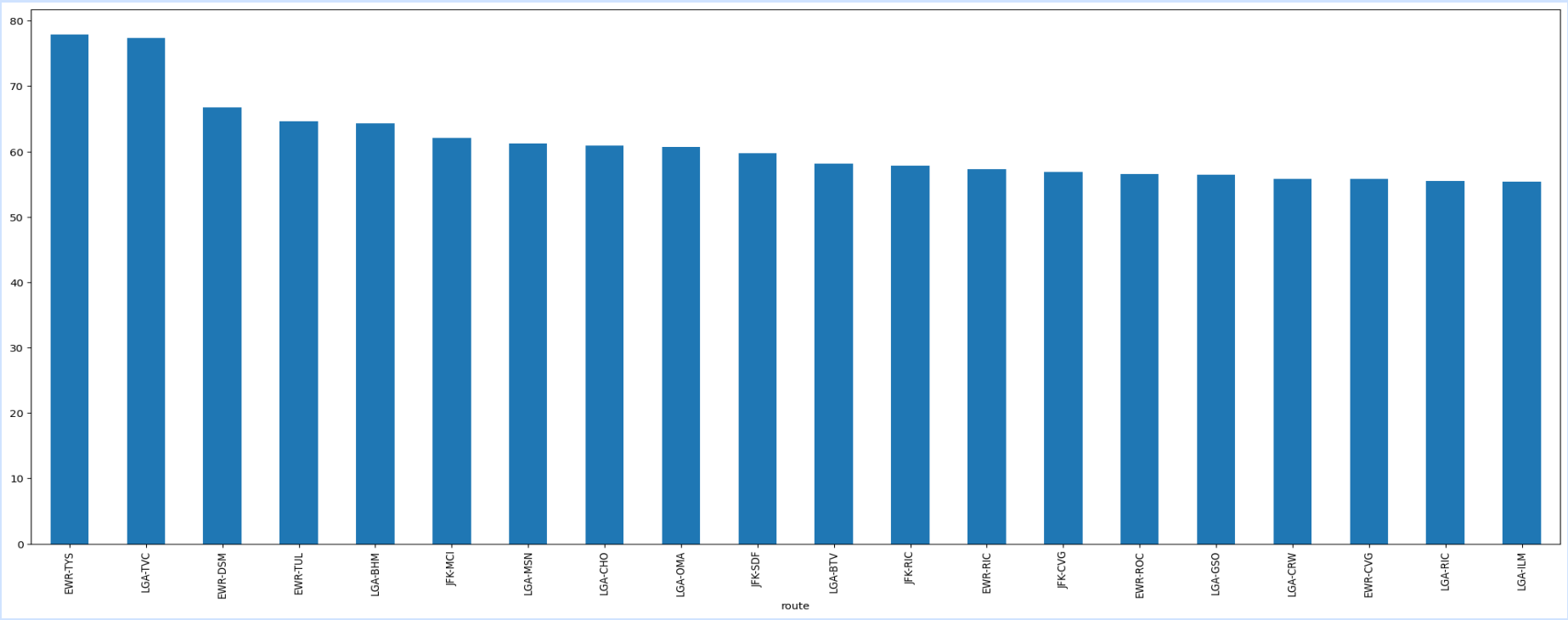
- Poor weather conditions such as heavy rain, snow, or strong winds leads to more fuel consumption in travel time.
- Inefficient route optimization planning, leading to longer travel time and delay in deliveries.
- Some of the airlines and carriers takes longer time than expected in making their deliveries to the customers.
- Lack of real time data analytics tools such as GPS tracking and real time data analytics for dynamic routes adjustments
- Traffic congestion is another factor that leads to high fuel consumption and travel time
- Poorly maintained fleet can lead to losses and damage of good as well as high fuel consumptions.

Kaggle Dataset: NYC airport routes Dataset analysis

- Below datasets analyses the factors that contributes to delays in different airlines at New York City



- High wind speed causes high fuel consumption, due to increased time for take off and landing.



- Upon analysis of Top 20 different routes, EWR-TYS was one of the routes with the highest delays, on average 78 minutes.
- These are the key areas DHL should focus on when making a choice of an airline to use to make deliveries.

Predictive analytic model for route optimization

- This involves the use of historical datasets and Machine learning techniques to create models that can be used to predict and optimize future routes by DHL. Models include:
  - Time series Analysis- Offers statistical features to analyze historical data an identify hidden trends.
  - Regression Models – Visualize the distribution of Data
  - Decision Trees – They are graphical models used in decision analysis
  - Neural Networks – They are artificial Intelligence model that mimic human brain.
- DHL can make use of predictive analytics models to make predictions on the best routes to be taken, which is a powerful tool to analyze big data and come up with a best performing mode.

Kaggle Datasets: Predictive Analytic Models

Linear Regression Model

-------------Classification Report-------------

	precision	recall	f1-score	support
0	0.76	0.97	0.85	41520
1	0.55	0.11	0.19	14318
accuracy			0.75	55838
macro avg	0.65	0.54	0.52	55838
weighted avg	0.71	0.75	0.68	55838

Random Forest Model

-------------Classification Report-------------

	precision	recall	f1-score	support
0	0.77	0.98	0.86	41520
1	0.70	0.14	0.23	14318
accuracy			0.76	55838
macro avg	0.73	0.56	0.54	55838
weighted avg	0.75	0.76	0.70	55838

XG Boot Model

-------------Classification Report-------------

	precision	recall	f1-score	support
0	0.81	0.95	0.87	41520
1	0.69	0.34	0.46	14318
accuracy			0.79	55838
macro avg	0.75	0.64	0.66	55838
weighted avg	0.78	0.79	0.77	55838

KNeighbor Model

-------------K Nearest Neighbors Results-------------

Accuracy: 0.7644256599448405

F1 Score: 0.4557716177079023 7

ROC AUC Score: 0.7091625390490324

Decision Tree Model

-------------Classification Report-------------

	precision	recall	f1-score	support
0	0.77	0.97	0.86	41520
1	0.63	0.17	0.26	14318
accuracy			0.76	55838
macro avg	0.70	0.57	0.56	55838
weighted avg	0.73	0.76	0.70	55838

Neural Network Model

Neural Network:

	precision	recall	f1-score	support
0	0.79	0.94	0.86	41520
1	0.64	0.29	0.40	14318
accuracy			0.78	55838
macro avg	0.72	0.62	0.63	55838
weighted avg	0.75	0.78	0.74	55838

Kaggle Dataset: Overall performance of the Model

Out[232]:

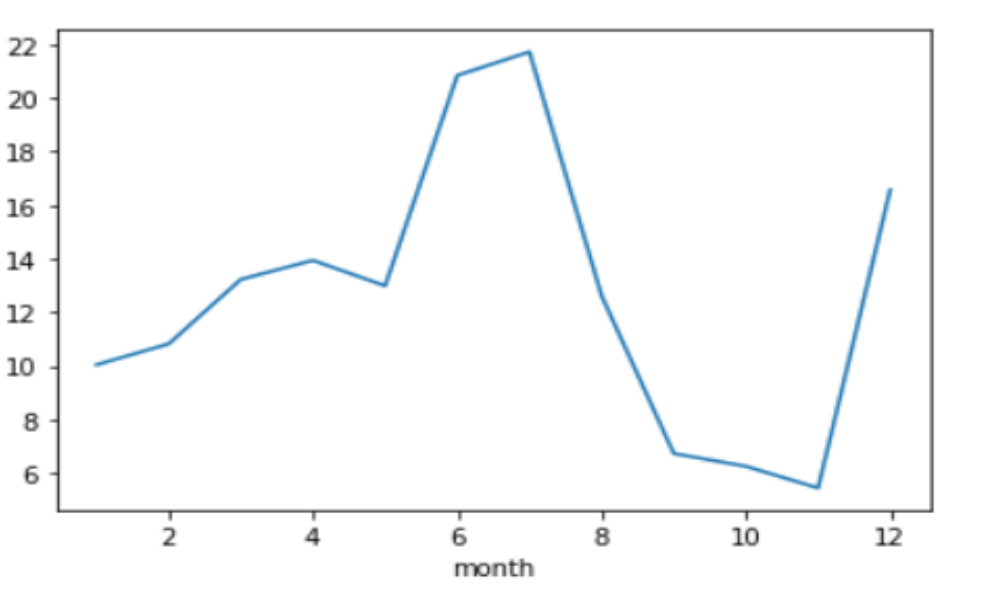
	Model Name	Accuracy	F1-score	ROC
1	LogisticRegression	0.748863	0.188061	0.694148
2	XG Boost	0.791361	0.458391	0.775728
3	Neural Network	0.776335	0.371486	0.746255
4	K Nearest Neighbors	0.764426	0.455772	0.709163
5	Random Forest	0.743580	0.000000	0.553440
6	Decision Tree	0.743580	0.000000	0.500000

- XG Boost model gives the best accuracy score of 79 % on prediction, followed by Neural network at 77.6%.
- When using ROC accuracy, XG Boost model also performs best at 77.5% followed by Neural Networks.
- These two models can be employed to evaluate the delay time when specific route is taken.
- DHL can employ the same techniques in their predictive and prescriptive analytics modelling.

Prescriptive analytic model for route optimization

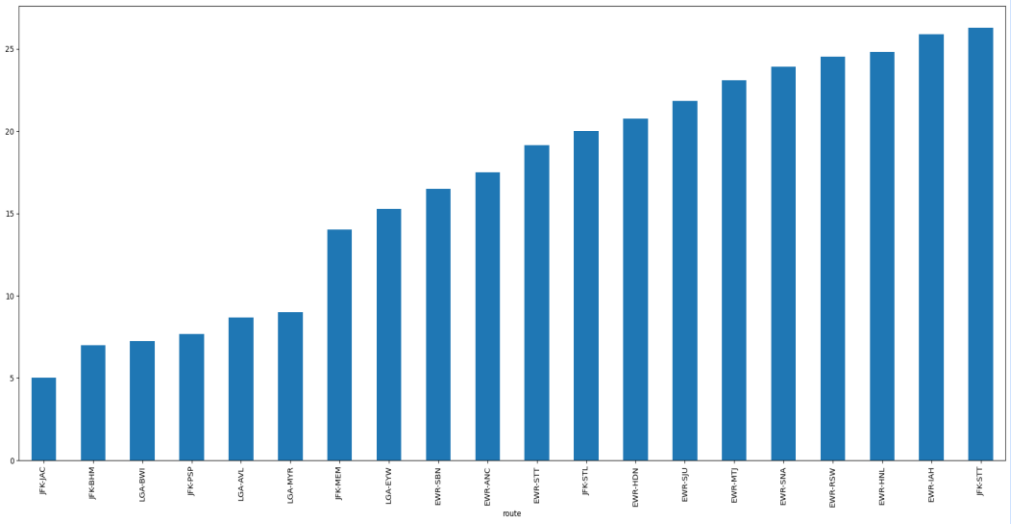
- These recommends the best actions to be taken on a model to achieve specific desired outcome.
- Suitable models include:
  - Decision models such as sensitivity analysis and decision tree.
  - Genetic algorithms,
  - Reinforcement learning for optimizing routes based on predictive insights.
  - Optimization models such as linear programing
  - Simulation models such as system dynamics and agent-based modeling
- These models can be employed and be improved on a day-by-day basis depending on hoe the flights and delays are.

Kaggle Dataset: Prescriptive insights.



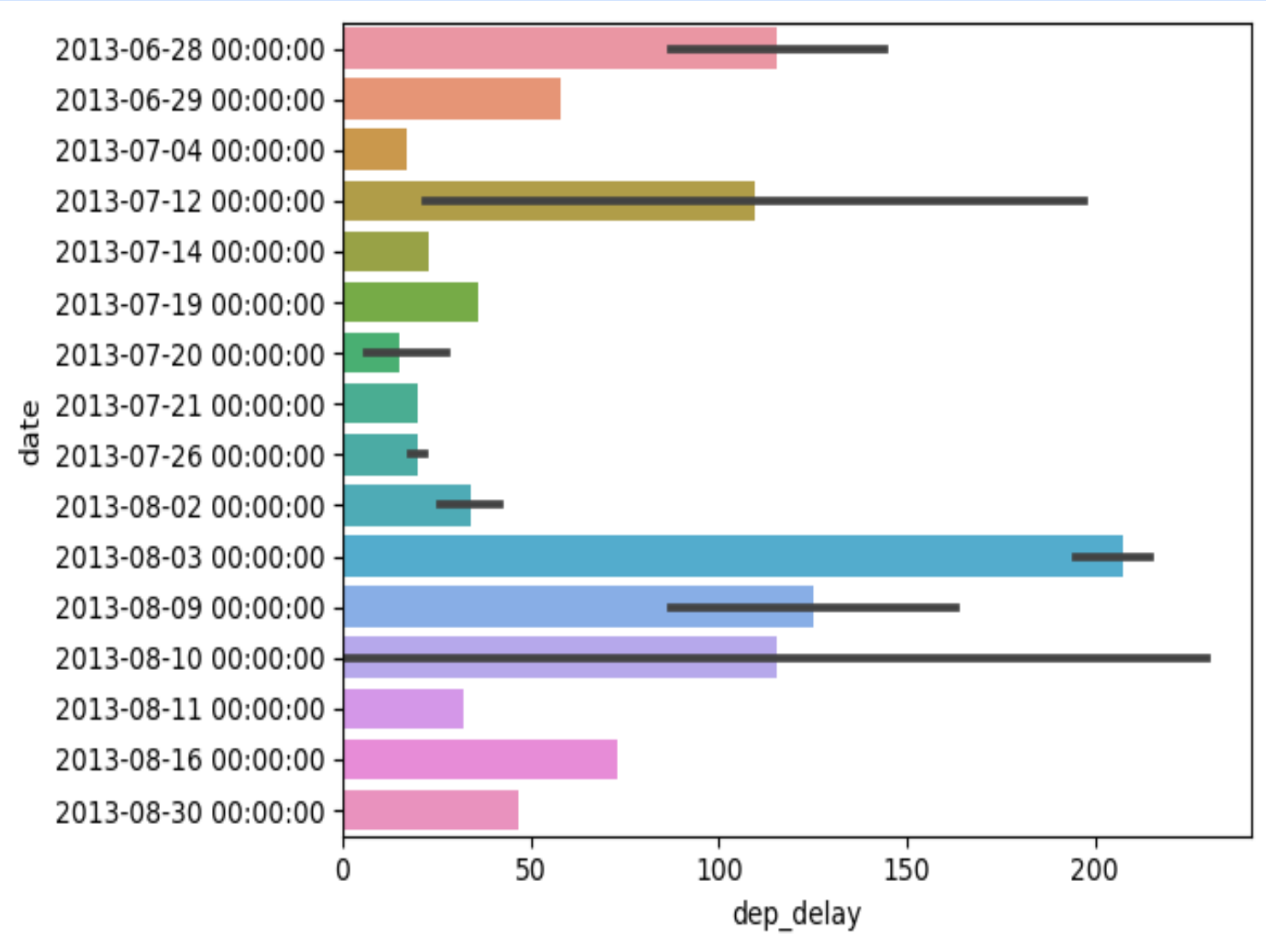
This is an analysis of number of flights delays in through the months. On average, the delay is high on month of June and July due to many as the flights is at peak season where bookings are very high.

This can be a powerful tool for DHL to optimize on delivery options during peak and off-peak seasons.



When reviewing most preferred airlines with less delays, JFK-JAC was had an average delay time of 5 minutes.
It is also good to review other factors such as the availability, cost and reliability of the airline when making deliveries.

Kaggle Dataset: Prescriptive insights.



This is a deeper analysis of the months with the highest number of delays on average.

EWR-TYS route was picked to dig deeper on the specific months which delay was the most and to evaluate on the best decision made based on the past data and the future predictions.

Implementation of prescriptive and predictive models at DHL

i. Define clear objectives on what the company needs to achieve and how data analytics can come in play.
ii. Gather relevant historical data on different routes, delivery times, traffic conditions, weather, vehicle maintenance, and other identified variables.
iii. Choose an appropriate machine learning algorithms to perform predictive analytics
iv. Formulate mathematical models for route optimization and select an optimization solver.
v. Integrate predictive and prescriptive models with existing operation systems
vi. Have a training plan to train personnel on analytics tools
vii. Ensure data governance and compliance are in place.
viii. Continue to monitor the models and improve when necessary

Quantifying and Measuring Goals

- Defining metrics: Fuel consumption rates, average travel time, on-time delivery rate, customer feedback analysis.
- Establishment of benchmarks for comparison.
- Tracking progress using key performance indicators (KPIs) such as customer satisfaction. KPIs should reflect the critical success factors for each objective.
- Implement real time data tracking and management systems, so that customers can track their goods.
- Evaluate performance of deployed models on the overall efficiency of the companies' operations.
- Communicate results to the team regularly on the target that have been hit and milestones accomplished.

Challenges and Limitations and Mitigation measures

- Data quality and availability- This can be resolved by Investing in data cleaning and preprocessing techniques.
- Model accuracy. Some models can give inaccurate results and that can ne resolved by regular model maintenance and updates.
- Formulating and solving complex problems may require advanced mathematical models and powerful solvers. This can be resolved by matching the complexity of the model to the resources available.
- Unpredictable conditions such as weather and traffic. This can be addressed by integrating real-time data and route optimization
- Scaling optimization models to handle large datasets and complex logistics networks can be challenging. This can be mitigated by leveraging parallel computing resources.

Cost Benefit Analysis

- Assessing the cost-effectiveness of implementing predictive and prescriptive analytics.
- Comparing the potential benefits (fuel savings, reduced travel time) against implementation costs of the analytics tools.
- Consider also the skills and expertise of the current staff before hiring an external
- Also provide a solid basis for comparing alternative solutions in data analytics tools that are not costly and they deliver best results, and they are compliant to Data Governance requirements.
- Establish the timelines for the established goals to be achieved in terms of revenue and cost.

Recommendations and next steps

- From the Flights Delay Classification on NYC Flights it is recommended to leverage on analytics tools to predict on the future and give the best possible solutions to the present state on the actions to be taken to achieve a desired results.
- It is also recommended for DHL to decide on the best route to be taken based on the cost, delays, weather, the type of airline, reliability among other factors to always ensure that customer needs are satisfied, and costs are saved.
- Train staff to always be data proficient and to make decisions based on insights gained from data analysis.
- It is recommended to leverage on machine learning and big data analytics to solve complex airlines problems and identify the best routes that safe costs

References

Omkar khandekar. 2020. Flights Delay Classification on NYC Flights data . Kaggle Dataset https://www.kaggle.com/code/omkarkhandekar/flights-delay-classification-on-nyc-flights-data/notebook