MGT 6203 Group Project Proposal Template

Team Information (1 point)

Team #: 13 – aka Definitely Vibrant

Team Members:

1. Spencer Allgaier; 903844347
   Full-time data scientist for the government with two years of experience, graduated from BYU in 2019 with a degree in Statistics and a minor in Business Management.

2. Eleibny F. Feliz Santana; 903626016
   Data Engineer with expertise in data structures design, development as well as ETL design, data transformation processes development and implementation. Previous experience performing Data Exploratory Analysis and developing dashboards to help drive leadership decisions.
   - BSc in Software Engineering - Universidad Iberoamericana (Santo Domingo, D.R.)
   - MSc in Distributed Systems – Heriot Watt University (Scotland, UK)

3. Thuy Thi Dinh; 902479253
   10 years' experience as a Mechanical Engineer in various industries including Aerospace, Biomedical Research and Manufacturing. Worked on various machine interfacing and query projects.
   - BS in Mechanical Engineering - Georgia Institute of Technology 2013

4. Leland Bolleter; 903747003
   Experience in DBA, SW development, and project management. MSCS and work with USDA on soil carbon capture.

Objective/Problem (5 points)

Project Title:
Optimization Input for Flight Schedules to Minimize Departure and Arrival Delays in Coordination with Risk Ranking of a Major US Airport Based on Past Flight Delays

Background Information on chosen project topic:

Approximately two million people use the US airports every day to reach domestic and international destinations. Air travel has become a popular mode of transport for a large portion of the world's population. Businesspeople want reliable schedules, families want on-time arrivals for vacations, and airlines want to provide on-time service to retain consumers.

Despite the airline's careful and meticulous planning, weather conditions at the time of departure is something that cannot be controlled. Predicting how and when different types of weather conditions correlate with flight delays can be helpful for both travelers and airlines to understand the delay potential and plan accordingly. Both airline companies, air travelers, and the collective airline industry in general strive to mitigate and reduce the waste inherent in weather-related-air-travel delays.

Problem Statement:
The purpose of this project is to correlate and predict weather-related-air-travel delays for flight departures and arrivals at one (or more) major US airport(s). Our goal is to use historical flight information and weather data to predict flight timeliness and related factors. With this knowledge it is intended that our customer will be able to reduce flight delay related waste and expenditures.

Primary Research Question (RQ):
How can past weather delays associated with an airport location or season be used to mitigate future flight delays?

Supporting Research Questions:

2. Are there particular airports or combinations of airports and seasons particularly prone to flight delays?
3. What recommendations can be made to reduce air-travel delays?
4. Which weather conditions are more likely to correlate with a delay?
5. Is there more impact to air-travel delays at the departure or the destination airport locations and/or seasons.

Business Justification:

The FAA estimated that in 2019 airlines incurred an 8.3 billion dollar loss due to flight delays, 2.5 billion of which can be attributed to inclement weather conditions. Passengers incurred a cost of 18.1 billion associated with time lost due to schedule buffers, delayed flights, flight cancellations, and missed connections. Also affected are ground crew, pilots, flight crew, and supporting industries that depend on air transportation for business operations. According to trefis.com, delays and cancellations are not only an inconvenience for passengers and airlines, but also result in significant financial consequences.

Predicting potential hot spots of weather-related-air-traffic delays would be important to improve the quality and timeliness of service provided by airlines. Improved service could be achieved by the development of a framework to facilitate an airline to hopefully avoid risky times and locations. The framework and predictions would also allow help with mitigation plans in situations where delays become probable based on recent activity. We envision that Markov Chains may help with these predictions.

Dataset/Plan For Data (4 Points)

The data used for this analysis is planned to be procured from 2 main resources:

- Flight information from the Bureau of Transportation Statistics, a resource supported by the U.S. Department of Transportation (DOT). Specifically, we plan to pull from the detailed statistics departure database which allows users to specify departure airport, airlines, and dates of interest.
- The weather data used in this project will be sourced from the National Centers for Environmental Information (NCEI). NCEI maintains one of the most significant archives on Earth and is supported by the National Oceanic and Atmospheric Administration (NOAA). We plan to pull the Daily Summaries from the Climate Database which allows specification of dates of interest along with location.

- We will use R and Excel to clean and prep the data for analytics.

Data Description

The flight departures data contain all the flight departures with delay details for a specified airport. The dataset is highly flexible and can be scaled to focus on key carriers with date ranges. The image below shows the first 10 lines of data for Delta Airlines at LaGuardia Airport during January 2022.

1	Carrier Code	Date	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time	Actual elapsed time	Departure delay	Wheels-off time	Taxi-Out time	Delay Carrier	Delay Weather	Delay National Aviation System	Delay Security	Delay Late Aircraft Arrival
2	DL	1/1/2022	414	N370DN	ATL	17:05	0:00	158	0	0	0:00	0	0	0	0	0	0
3	DL	1/1/2022	433	N338DN	ATL	10:46	10:45	169	143	-1	11:05	20	0	0	0	0	0
4	DL	1/1/2022	447	N367DN	ATL	6:00	0:00	154	0	0	0:00	0	0	0	0	0	0
5	DL	1/1/2022	451	N305DU	ORD	9:28	0:00	160	0	0	0:00	0	0	0	0	0	0
6	DL	1/1/2022	490	N340DN	ATL	15:59	16:15	162	140	16	16:31	16	0	0	0	0	0
7	DL	1/1/2022	492	N358DN	ATL	18:45	19:39	162	137	54	19:54	15	29	0	0	0	0
8	DL	1/1/2022	507	N105DX	ATL	14:45	15:14	157	137	29	15:26	12	0	0	0	0	0
9	DL	1/1/2022	529	N313DN	ATL	12:42	14:59	160	141	137	15:21	22	118	0	0	0	0
10	DL	1/1/2022	531	N310DU	DFW	14:00	13:56	259	264	-4	14:12	16	0	0	0	0	0

The weather data contains daily weather information for a given city. It details numeric weather information of snowfall, precipitation, wind speed, and additional factors. The weather data also provides categorical information for weather types which cannot be easily measured but still impact flight logistics.

1	Station	Name	Date	Average Wind	Precipitation	Snowfall	Snow Depth	Average Temp	Max Temp	Min Temp	WT01 (Fog)	WT02(Heavy Fog)	WT03 (Thunder)	WT04(Sleet)	WT05(Hail)	WT08(Smoke)	WT11(High Damaging Winds)
2	USW00014732	LAGUARDIA AIRPORT, NY US	1/1/2022	6.26	0.76	0	0	53	57	50	1						
3	USW00014732	LAGUARDIA AIRPORT, NY US	1/2/2022	9.62	0.04	0	0	53	60	39	1						
4	USW00014732	LAGUARDIA AIRPORT, NY US	1/3/2022	14.32	0	0	0	35	39	24							
5	USW00014732	LAGUARDIA AIRPORT, NY US	1/4/2022	8.72	0	0	0	27	36	21							
6	USW00014732	LAGUARDIA AIRPORT, NY US	1/5/2022	7.61	0.24	0	0	38	48	32	1					1	
7	USW00014732	LAGUARDIA AIRPORT, NY US	1/6/2022	11.18	0	0	0	41	43	36							
8	USW00014732	LAGUARDIA AIRPORT, NY US	1/7/2022	18.57	0.33	9.7	7.9	33	36	28	1						
9	USW00014732	LAGUARDIA AIRPORT, NY US	1/8/2022	10.74	0	0	5.1	27	32	23							
10	USW00014732	LAGUARDIA AIRPORT, NY US	1/9/2022	10.96	0.11	0	3.1	35	42	30							

Key Variables:

Our dependent variable will be if a weather delay was found. We do not intend on creating new variables, although the weather categorical data will be transformed into dummy variables. The focus will be based on the impact of weather factors so we hypothesize that rain, wind speed and weather type will have the biggest impact on the dependent variable.

Approach/Methodology (8 Points)

We will obtain historical weather data and flights departure and arrival data, including delays for a major US airport. The data will be processed and cleaned by removing duplications, filtering any unwanted outliers, decide how to handle missing information, and aggregating and transforming the data to the granularity needed.

Initial investigations will be performed by using Exploratory Data Analysis to spot anomalies, identify patterns and to check assumptions. In our approach, we'll focus on highly-interpretable-linear-regression models for preliminary results. We also anticipate that some more flexible but less interpretable models such as deep learning or GAMs will be useful to explore. To compare models, we plan to use a 80/20 train and validation segmentation of the data. We also plan to use cross-validation for a full set of training, validation, and test data.

Then, we'll explore the results to learn about what factors are correlated with delays. We anticipate interesting information may be revealed with factors such as hub location, thunderstorms, Jetstream, time-of-day, and more. We are still in the investigatory phase of data gathering and wrangling. We do not have examples of data transformations currently.

Some of the techniques we plan to use for visualizations are: *Scatter plots:* which are useful to show the relationship between the response and the predictors. *Line graphs*, to plot continuous data and see trends and *Bars.* These provide a powerful and easy to understand media to communicate results. We'll also investigate the use of Tableau and the vast array of graphics including geospatial visualizations.

Anticipated Conclusions/Hypothesis:

We expect to see significant delays correlating with thunderstorms and snowstorms. We plan to investigate and possibly find a point in severity where there is an inflection in quantity of delays. We also expect to see a decreasing number of flights leaving on time with the increase in severity of weather.

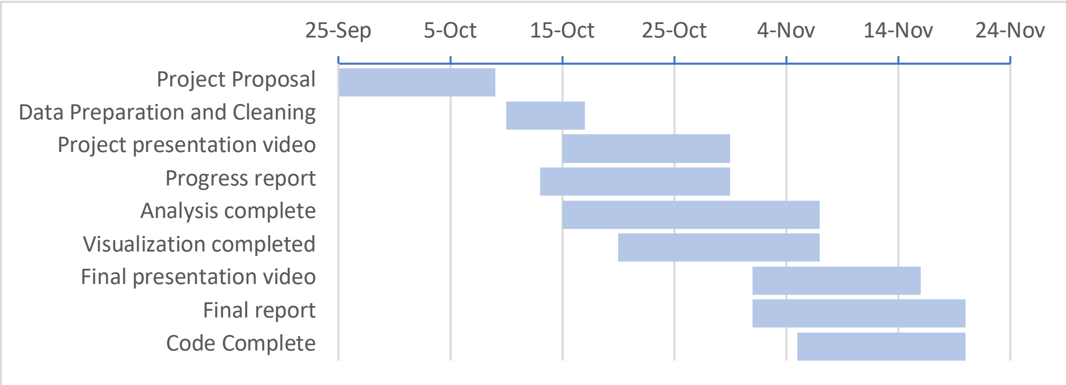
What business decisions will be impacted by the results of your analysis? What could be some benefits?

By understanding the weather impact on flight schedules, individuals and businesses can make better decisions to avoid delays when booking travel. Research shows that even though business travelers make up only 12% of airline passengers, they make up 75% of total profits. It is imperative for airlines to retain business customer loyalty by providing confidence of service. Building a reputation for reliability, convenience and timeliness will allow airlines to grow their most valuable customer base.

Airlines could further leverage this projects predictive model to calculate delay risk for flights. The risk calculation could help to avoid paying out customer compensation from rebooking, refunds and overnight accommodations. It would also reduce the risk of flight cancelations due to compounding delays from multiple segments planes typically fly in the continental US per day. To accomplish this, airlines can use this projects model(s) to preemptively exclude or rearrange times for at risk segments.

Project Timeline/Planning (2 Points)

Project Gantt Chart



October 9 - Proposal complete - Spencer Allgaier (lead) with Team assistance
October 17 - Data prepped and cleaned - Thuy Dinh (lead) with Team assistance
October 30 - Project plan presentation video - Eleibny Feliz Santana (lead) with Team assistance
October 30 - Progress report complete - Leland Bolleter (lead) with Team assistance
November 7 - Analysis complete - Eleibny Feliz Santana (lead) with Team assistance
November 7 - Visuals complete - Spencer Allgaier (lead) with Team assistance
November 16 - Final presentation video complete –Thuy Dinh (lead) with Team assistance
November 20 - Final report - Leland Bolleter (lead) with Team assistance
November 20 - Code complete- Spencer Allgaier (lead) with Team assistance

Appendix (any preliminary figures or charts that you would like to include):