Analytics Startup Plan

Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.*

Project	*Predicting Airline Passenger Satisfaction*
Requestor	*Centennial College*
Date of Request	*July 10, 2023*
Target Quarter for Delivery	*August 18, 2023*
Epic Link(s)	*July 11, 2023*
Business Impact	*Predicting the satisfaction of passengers will help airline companies to predict which factors lead to passenger satisfaction or dissatisfaction. Once airlines get to know about these factors, then they can work on them to make their passengers more satisfied with their services. This will aid the companies to retain their customers in such a dynamic environment and hence boost their sales.*

1.0 Business Opportunity Brief

Clearly articulated business statement of the Ask, opportunity, or problem you are trying to solve. An important step is to understand the nature of the business, system, or process and the desired problems to be addressed. This will be communicated back to All stakeholders for alignment.

The objective of this capstone project is to predict airline passenger satisfaction by leveraging business analytics techniques. Airlines may make data-driven decisions to optimize their operations, service offerings, and customer interactions by analyzing and anticipating passenger satisfaction, ultimately leading to enhanced customer loyalty and financial success. Passenger satisfaction is a significant aspect of defining an airline's reputation and performance in a highly competitive sector. The current difficulty stems from a lack of thorough understanding of the aspects that have a substantial impact on passenger satisfaction. While airlines collect massive amounts of data on customer feedback, preferences, demographics, flight details, and other pertinent information, significant insights must be extracted from this data to discover the major drivers of satisfaction.

The specific ask:

Clearly articulate the specific task you will be conducting to help achieve the opportunity.

1. To develop the model and analyze its results to identify significant factors influencing passenger satisfaction.
2. Extract meaningful insights from the findings and provide practical suggestions to airlines to enhance customer experiences and improve overall service quality.

1.1 Supporting Insights

Define any supporting insights, trends, and research findings. Where relevant, list key competitors in the market. What are their key messages, products & services? What is their share of the market, nationally and regionally?

According to the dataset, some of the following aspects influence passenger satisfaction: Inflight Wi-Fi service, Ease of Online booking, Food and drink, Online boarding, Inflight entertainment, on-time performance, baggage handling, in-flight services, etcetera. The airline sector is undergoing tremendous digital change. Airlines are investing in technology to improve the customer experience, such as ticket booking apps, self-check-in kiosks, real-time flight updates, and personalized recommendations based on passenger preferences. The rising emphasis on environmental sustainability is one notable trend in the aviation sector. Airlines are recognizing the need to combat climate change and reduce their carbon footprint. This trend is being driven by numerous reasons, including Climate change awareness, regulatory pressure, cost savings, etcetera.

1.2 Project Gains

Airlines may make data-driven decisions to improve the customer experience by accurately predicting airline passenger happiness, leading to improved customer loyalty and possible revenue advantages. Customers who are satisfied are more likely to return to the airline, promote it to others, and potentially spend more on ancillary services and upgrades. Understanding the elements that have the greatest influence on passenger satisfaction allows airlines to discover areas for quality improvement. By resolving these issues, airlines can improve overall service quality, resulting in better customer experiences, favorable word-of-mouth, and a stronger market reputation. Using a predictive model for passenger satisfaction will allow airlines to handle customer requirements and preferences more proactively, resulting in a more personalized and fulfilling travel experience. Customers would value airlines that understand and cater to their specific interests, resulting in increased satisfaction and loyalty.

Note: Completion of the following sections is possible only after a careful assessment and triage of the Ask. This is required to determine scope, resource, time, priority, and data availability.

2.0 Analytics Objective

The primary objective of this research is to find the best model with the highest accuracy rate for predicting passenger satisfaction.

Moreover, this research will help to answer the following questions:

1. How well can we predict passenger satisfaction based on historical data of customer feedback and reviews?
2. Can we predict passenger satisfaction levels based on a combination of flight-related factors (e.g., ease of online booking, online boarding, check-in services) and demographics-related factors (e.g., age, type of travel, gender)?
3. What are the key factors that significantly impact airline passenger satisfaction?

4. What are the primary reasons behind customer complaints or negative feedback regarding airline services?
5. What amenities or services do passengers value the most in terms of enhancing their satisfaction?

2.1 Other related questions and Assumptions:

> ℹ️ *List any assumptions that may affect the analysis*

Assumptions that may affect the analysis:

1. It is assumed that the 'Flight distance' column is measured in kilometers.
2. In the dataset, it is assumed that if a flight had "First class", it is mentioned as "Business" in the 'Class' column.
3. Satisfaction levels are given in the range of 0-5 where 0 is defined as not applicable and 1 is defined as the lowest satisfactory level and 5 is the highest.
4. It is assumed that all delays either in arrival or departure beyond 30 minutes are capped.
5. Time of delay in the 'Departure delay' and 'Arrival delay' columns is assumed to be delayed by the actions of airlines and not by any external factors.
6. Feedback provided by passengers is a rationale.
7. All the information provided in the dataset is legitimate.
8. The feedback provided by passengers is true.
9. The feedback by children is provided under their parent's supervision and is true.

2.2 Success measures/metrics

> *What does success look like? Define the key performance indicators (success definition/indicators, drivers, and key metrics) against which the objectives will be analyzed. These should be drawn from the interlock meeting with key stakeholders and will inform the approach and methodology for the analysis.*

Key performance indicators for this research are as follows:
1. Increase in customer retention rate for the companies.
2. Trust of the passengers will be increased if their feedback is valued.
3. Increase in profits for the company.
4. Determine the significance and impact of various variables on passenger satisfaction.
5. Assess the usefulness and applicability of the analysis's insights in boosting customer satisfaction.
6. The predictive accuracy of the model will be measured by root mean square error.

2.3 Methodology and Approach

Now that you have a good understanding of the Ask and deliverable, detail the recommended approach/methodology.

Type of Analysis: Decision tree with Grid Search, Random Forest, Regression models, and Gradient Boosting.

Initially, I will be creating Decision Trees to analyze feature importance and which features are more significant. I will later deploy other techniques to verify my findings.

Methodology: Key assumptions and questions from 'Analytics objective' will be considered and analysis will be performed as outlined in the '5.0 Timelines and deliverable section'.

I will start with Exploratory Data Analysis which will provide more understanding of the data and make it clear and clean as per my needs. Further, I will deploy feature engineering such as One-Hot Encoding, Feature Transformation, etcetera as per my needs. Moreover, I will be building models as mentioned above and then selecting the best model based on accuracy score and ROC-AUC score.

Output: The output of this result will be the classification of passenger satisfaction. This output will then be analyzed to get meaningful insights which will be recommended to the companies to create their strategies accordingly and achieve the KPIs mentioned in 2.2.

3.0 Population, Variable Selection, considerations

Capture learning about the data available today location, structure, and reliability; this would include data in operational systems including dealer sourced, data warehouse and any CRM or email marketing systems available today.

Audience/population selection: Vice President of an Airline Company

Observation window: Not provided.

Inclusions: Airlines Passenger Satisfaction Survey data.

Exclusions: **'Unnamed: 0'** and **'id'** columns are excluded from the dataset for this research.

Data Sources: https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction?select=train.csv

Audience Level:

Variable Selection: Selecting **'satisfaction'** as our target variable.

Derived Variables: Binning variables as per needs.

Creating Dummy variable.

Converting numerical variables to categorical where necessary.

Assumptions and data limitations: Assumptions are as follows:
1. It is assumed that the 'Flight distance' column is measured in kilometers.

2. In the dataset, it is assumed that if a flight had "First class", it is mentioned as "Business" in the 'Class' column.
3. Satisfaction levels are given in the range of 0-5 where 0 is defined as not applicable and 1 is defined as the lowest satisfactory level and 5 is the highest.
4. Time of delay in the 'Departure delay' and 'Arrival delay' columns is assumed to be delayed by the actions of airlines and not by any external factors.

4.0 Dependencies and Risks

ℹ️ *Identification of key factors that may influence the outcome of the project and the likelihood of it happening:*

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
1. *There are three unique values in the satisfaction column but 'neutral or dissatisfied' is grouped as one and 'satisfied' as the other.*	*Low*	*2-3 days*	*It will be difficult for the Airline companies to know which passenger is dissatisfied with their services. Identifying dissatisfied passengers and effectively addressing their concerns will be a crucial challenge for them.*
2. *Delay in peer reviewing.*	*Low*	*1-2 days*	*This may impact the delay in necessary revisions recommended by peers.*

5.0 Deliverable Timelines

ℹ️ *List key dates and timelines as a work-back schedule. Activate line items based on complexity and line-of-sight required. Will set the stakeholder expectations for the process.*

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	*Kick-off/ Analysis Plan*	*Building Analysis Plan*	*Pen down the blueprint*	*7*	*July 11, 2023*
2.	*Exploratory Data Analysis (EDA)*	*Describing the data and the approach to follow, summarizing the data*	*Providing brief descriptive statistics,*	*7*	*July 18, 2023*

		features, and cleaning the data	*Transforming the variables,* *Imputing the values*		
3.	*Modeling*	*Building various models to predict passenger satisfaction*	*Building models like Decision Trees, Random Forests, Linear Regression, and Neural Networks*	*8*	*July 24, 2023*
4.	*Report Building*	*Build a formal report which will be submitted for grading*	*Start on building the report and defining each and everything starting from the problem statement and ending with recommendation s*	*7*	*August 1, 2023*
5.	*Gov & Documentation*	*Documenting the report as per standards*	*Thorough documentation of data sources, collection methods, and any data transformations needed. It involves proper storytelling.*	*5*	*August 8, 2023*
6.	*Peer Review*	*Review by peers*	*The report will be reviewed by peers and feedback would be provided*	*3*	*August 13, 2023*
7.	*Presentation*	*Presenting the report in front of the stakeholder.*	*The final presentation will be presented in front of the various stakeholders with disclosures of all relevant facts, according to which the report will be graded.*	*2*	*August 16, 2023*

8.	Submitting the final work	Assignment and portfolio submission.	The report and the presentation will be submitted for grading.	1	August 18, 2023