**Executive Summary: Enhancing Airfare Decision-Making Through Data Analysis**

Commercial airfare has become the cornerstone of long-distance travel in the United States, marking a paradigm shift in transportation norms. However, as airlines evolve their systems and processes to optimize revenue, transparency for consumers has waned. The proliferation of hidden costs, including delayed or canceled flights, baggage charges, employee dissatisfaction, and maintenance issues, has clouded the true cost of air travel. This project delves into the multifaceted realm of airline operations, aiming to uncover the true cost of flights while proposing a risk-cost-based optimization model to empower customers in making airfare decisions.

1. Problem Description:

Key Performance Indicators (KPIs) in the airline industry encompass a broad spectrum, ranging from accident rates to maintenance efficiency. This project recognizes the pressing need to unravel the complexities of airline operations to empower consumers in their airfare decisions. By delving into the myriad metrics that define airline performance, we aim to shed light on the true cost implications and enable customers to navigate the airfare landscape with confidence.

2. Data Curation:

Data acquisition involved sourcing information from diverse platforms such as the Bureau of Transportation Statistics, Kaggle, IBM, SerpApi’s Google Flight API, and MIT. The datasets comprised a plethora of variables including baggage data, passenger-to-employee ratios, delay statistics, and airfare information. Most of the data could be downloaded as a CSV file. We obtained data from SerpApi’s Google Flight API and MIT through web scraping in Python to create data frames that could be saved as CSVs.

3. Data Cleaning:

While the richness of airline data facilitated analysis, cleaning was crucial for ensuring accuracy. We employed techniques such as handling NA values, mitigating outliers, and standardizing airline names to enhance data quality. To maximize results and data obtained, we groomed the data for years 2009 to 2019. For most of the data, we could only compile complete datasets for major airlines and had to further trim these airlines due to airline company mergers. We also chose to narrow the data object types to standardize the data even more.

4. Data Structuring:

Normalization was undertaken to address scale disparities among airlines, facilitating meaningful comparisons. Additionally, regional and state categorizations were introduced to provide contextual insights.

5. Data Analysis:

The data analysis phase of the project involved comprehensive exploration and examination of the curated datasets. Key aspects analyzed included flight delays, baggage handling records, net income trends, and passenger-employee ratios across major airlines. Using advanced analytical techniques and visualization tool in Pandas\_GUI we gained insights into operational efficiency, financial health, and employee productivity within the airline industry. Initial findings suggest significant variations among airlines in managing delays and baggage handling, with Delta, Alaska, and Allegiant emerging as strong performers across multiple metrics.

6. Results Visualization:

Our visualization efforts aimed to present the complex airline data in a digestible and interpretable format. Through graphical representations and interactive visualizations generated using Pandas\_GUI, we conveyed key insights regarding airline performance and operational dynamics. A single visualization tool was able to create code-free plots in seconds that are similar quality to those produced in Ploty and Seaborn. Visualizations included trend analysis of delay incidents, comparison of net income trends among airlines, and exploration of passenger-employee ratios over time. These visualizations not only enhanced our understanding of the data but also facilitated communication of findings to stakeholders.

7. Predictive Modeling:

While the project has laid a solid foundation in data analysis and visualization, future endeavors will focus on predictive modeling to forecast airline performance and optimize flight selection processes. Leveraging predictive analytics techniques, we developed a model capable of assessing the financial impact of operational inefficiencies. By integrating predictive modeling into our analytical framework, we anticipate enhancing the decision-making capabilities of consumers and stakeholders in the aviation industry.

Conclusion:

This project endeavors to demystify the opaque world of airline operations, enabling customers to make more informed airfare decisions. By synthesizing and analyzing diverse datasets, we have started to develop a model to understand flight costs and risks. Moving forward, the development of an interactive tool or web application holds promise for personalized risk assessments and tiered flight recommendations, further empowering consumers in their air travel endeavors.