



APAN 5900
SOLVING REAL WORLD PROBLEMS
W/ ANALYTICS

FLYAHEAD CONSULTANTS

PROJECT PROCESS FLOW



A 101 Handbook for our Capstone Project



TABLE OF CONTENTS

1. Introduction.....3

2. Project Process Flow

 a. Problem Identification.....3

 b. Data Inspection.....3

 c. Data Cleaning.....4

 d. Selecting Relevant Data for Analysis (Leisure Travelers).....4

 e. Data Visualization and Insights.....4

 f. Feature Construction.....5

 g. Feature Selection.....5

 h. Modeling.....5

 i. Standardization / Scaling.....6

 j. Choosing the Optimal K Value.....6

 k. Choosing the Right Modeling Technique.....6

 l. Analyzing Distribution Amongst Clusters.....7

 m. Developing Recommendations and Strategies.....7

 n. Deliverables.....7

 o. Challenges and Mitigation.....8

3. Process Flow Chart.....9

4. Conclusion.....10

Driving
Innovation,
Elevating
Travel
Experiences

Emphasizes the
company's
dedication to
fostering
collaboration,
innovation, and
customer focus to
deliver excellence in
travel and driving
sustainable growth

Presented to:
American Airlines

AMERICAN AIRLINES CUSTOMER SEGMENTATION PROJECT



Introduction

This document provides a detailed overview of the structured and data-driven methodology employed for the American Airlines customer segmentation project. Each step was designed to address specific challenges in understanding leisure travelers' behaviors, preferences, and revenue potential. By leveraging advanced analytical techniques and iterative modeling, the project delivers actionable insights to enhance marketing strategies and operational efficiency. This process flow highlights the tasks, methods, and deliverables that collectively enabled the development of tailored recommendations to support American Airlines' strategic goals.

Project Process Flow

1. Problem Identification

- **Objective:** Enhance leisure traveler segmentation for American Airlines to improve targeted marketing and revenue generation.
- **Key Challenges:**
 - Traditional segmentation approaches failed to capture nuanced customer behaviors.
 - Complexity in travel booking patterns driven by travel purposes, booking windows, and seasonality.

2. Data Inspection

- **Objective:** Understand the dataset structure, quality, and variables to identify initial cleaning and transformation needs.
- **Steps Taken:**
 - Reviewed missing values and outliers.
 - Assessed data types and consistency across features.
 - Evaluated the relevance of variables for customer segmentation.
- **Code Reference:** Refer to code lines [14 – 17] in *Preprocessing and EDA.ipynb* where missing values, data types, and variable checks are performed.



3. Data Cleaning

- **Objective:** Ensure the dataset is error-free and ready for analysis.
- **Steps Taken:**
 - Removed invalid data entries (e.g., negative booking dates).
 - Addressed missing values using imputation where necessary.
 - Standardized date formats and categorical labels.
- **Code Reference:** *Refer to code lines [18 – 35] in [Preprocessing and EDA.ipynb](#) where invalid entries are handled and formats are standardized.*

4. Selecting Relevant Data for Analysis (Leisure Travelers)

- **Objective:** Narrow the dataset to focus on leisure travelers for meaningful segmentation.
- **Steps Taken:**
 - Filtered travelers using travel probability columns for Vacation, Personal, and VFR categories.
 - Excluded business travelers and low-probability leisure segments.
- **Code Reference:** *Refer to filtering steps, code line [36], in [Preprocessing and EDA.ipynb](#) where leisure travelers are selected.*

5. Data Visualization and Insights

- **Objective:** Use visualizations to uncover trends, patterns, and potential clusters in the dataset.
- **Steps Taken:**
 - Generated histograms for advanced purchase distributions.
 - Created density plots for revenue and WTP-related features.
 - Used correlation heatmaps to identify relationships between variables.
- **Code Reference:** *Refer to visualization steps, code lines [37 – 81], in [Preprocessing and EDA.ipynb](#), where histograms, density plots, and heatmaps are created.*



6. Feature Construction

- **Objective:** Engineer new features to capture customer behaviors and Willingness to Pay (WTP).
- **Constructed Features:**
 - Revenue Per Mile (RPM): Insight into price sensitivity relative to distance.
 - Revenue/AP Ratio: Indicates urgency and high WTP for late bookings.
 - Cumulative Advanced Purchase Ratio: Tracks booking patterns over time.
 - Distance-Adjusted Revenue: Captures premium payments relative to norms for travel distances.
- **Code Reference:** *Locate feature engineering steps, code lines [20 – 25], in `Variable_Importance.ipynb`.*

7. Feature Selection

- **Objective:** Select the most relevant features for clustering by analyzing their importance and correlations.
- **Steps Taken:**
 - Analyzed feature correlations and their relevance to segmentation.
 - Selected the following features based on their significance:
 - Revenue/AP Ratio, RPM, Distance-Adjusted Revenue, Weekend and Holiday Interactions, and Channel_RPM_Ratio.
- **Code Reference:** *Refer to code lines [57 – 60] for defining features and target variables and code lines [61 – 63] for the output for Variable Importance in `Variable_Importance.ipynb`.*

8. Modeling

- **Objective:** Implement clustering to segment leisure travelers based on behavior and revenue potential.
- **Steps Taken:**
 - Ran unweighted K-means for a baseline model.
 - Used weighted K-means to amplify revenue and WTP-related



features.

- **Code Reference:** Refer to code lines [39 – 45] in *K-Means.ipynb* and [43 – 49] in *Weighted K-Means.ipynb* for model implementation.

9. Standardization / Scaling

- **Objective:** Normalize data to ensure equal weight across features during modeling.
- **Steps Taken:**
 - Standardized continuous features (e.g., RPM, Revenue/AP Ratio) to avoid scale bias.
 - Ensured all features had mean = 0 and standard deviation = 1 for consistent clustering.
- **Code Reference:** Check data scaling steps, code lines [43 – 44] in *K-Means.ipynb* and code lines [47 – 48] in *Weighted K-Means.ipynb*.

10. Choosing the Optimal K Value

- **Objective:** Determine the ideal number of clusters.
- **Steps Taken:**
 - Applied the Elbow Method to identify diminishing returns on inertia reduction.
 - Conducted silhouette analysis to validate cluster separability.
 - Selected 5 clusters as optimal.
- **Code Reference:** Locate elbow method, code line [46] in *K-Means.ipynb* and code line [50] in *Weighted K-Means.ipynb*.

11. Choosing the Right Modeling Technique

- **Objective:** Determine the best clustering approach to capture nuanced behaviors.
- **Steps Taken:**
 - Proposed K-prototype clustering to handle mixed data types (not feasible due to technical limitations).
 - Pivoted to K-means and weighted K-means for effective segmentation.

Code Reference: Refer to code lines [53 – 61] in *K-Means.ipynb* and [57– 69] in *Weighted K-Means.ipynb* for cluster visualization and slide 19 – 27 in *American Airlines.pdf* for model comparison.



12. Analyzing Distribution Amongst Clusters

- **Objective:** Understand the characteristics and revenue potential of each cluster.
- **Steps Taken:**
 - Profiled clusters based on revenue, booking behaviors, and travel preferences.
 - Identified key clusters with high revenue contributions (e.g., Clusters 2 and 4).
- **Code Reference:** *Refer to cluster profiling steps, code lines [53 – 61] in [K-Means.ipynb](#) and [57 – 69] in [Weighted K-Means.ipynb](#) and the analysis was put in slides the [American Airlines.pdf](#). Slides 31-36 talk about the cluster profiles and slide 37 talk about the recommendations.*

13. Developing Recommendations and Strategies

- **Objective:** Translate cluster insights into actionable recommendations for marketing and operations.
- **Steps Taken:**
 - Designed strategies for revenue maximization (premium services for Clusters 2 and 4).
 - Proposed cost-efficient campaigns for low-revenue clusters (Clusters 1 and 3).
 - Developed channel-specific recommendations to target segments effectively.
- **Code Reference:** *Highlighted relevant insights derived from cluster analysis in [Weighted K-Means.ipynb](#) and [K-Means.ipynb](#) in the [American Airlines.pdf](#) in slides 22 – 30.*

14. Deliverables

- **Final Presentation:** Showcased findings and recommendations.
- **Results Document:** Detailed analysis and insights.
- **Dataset:** Cleaned dataset with defined features.
- **Scripts:** Python code with comments for reproducibility.
- **Process Flow Document:** Written explanation of steps, mirroring this document.



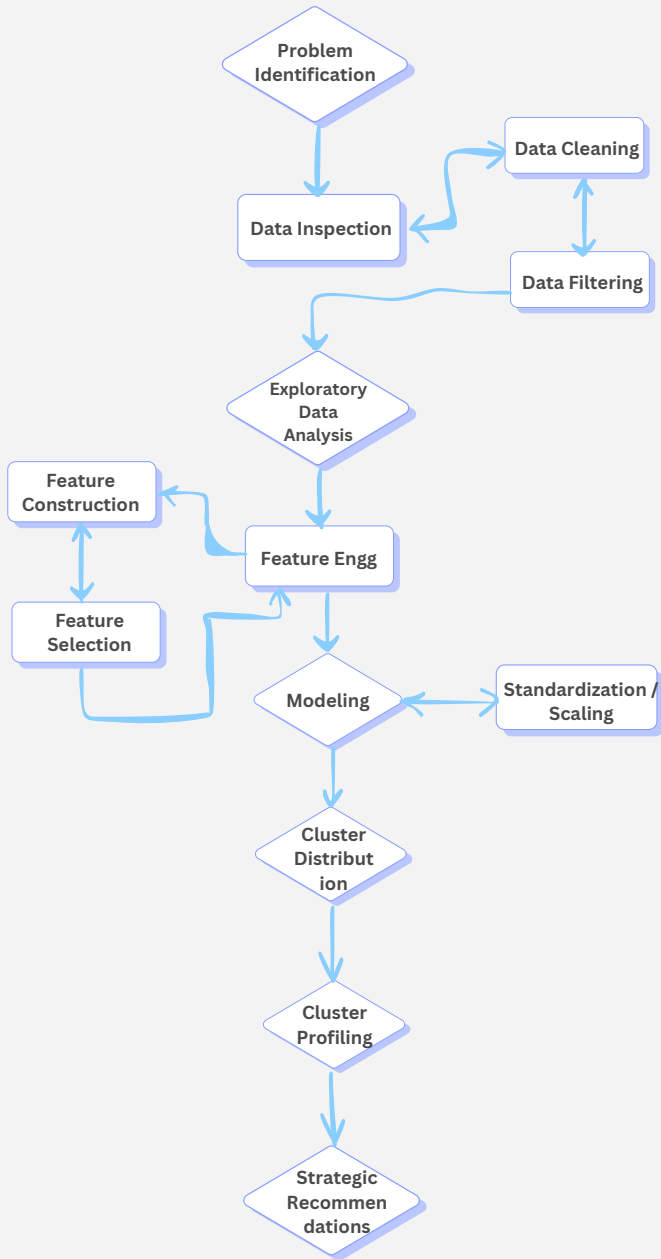
- **List of file names:** *These are all the files: **Preprocessing and EDA.ipynb**, **Variable_Importance.ipynb**, **K-Means.ipynb**, **Weighted K-Means.ipynb**, **Results and Recommendations**, **American Airlines.pdf**, **APAN 5900_Project Process Flow.pdf**, **bookingsData.csv**, **Distance_of_All_Airports.csv**, **Sample100_Data.csv**.*

15. Challenges and Mitigation

- **K-Prototype Clustering Failure:** *Memory limitations led to a pivot to K-means approaches.*
- **High Dimensionality:** *Addressed through feature selection and normalization.*
- **Dynamic Market Conditions:** *Acknowledged the need for iterative updates to maintain relevance.*
- **Code Reference:** *Documented issues and pivots in iterative **Sponsor Updates** and **Results and Recommendations** documents.*



Process Flow Chart





Conclusion

The American Airlines customer segmentation project demonstrates the power of data-driven strategies to address complex business challenges. By leveraging advanced clustering techniques, we identified five distinct traveler segments, providing actionable insights to optimize marketing efforts, enhance operational efficiency, and maximize revenue.

This comprehensive analysis not only highlights the diversity of leisure traveler behaviors but also equips American Airlines with tailored recommendations to engage high-value customers, efficiently manage resources, and sustain competitive advantage. The deliverables, including the final dataset, scripts, and process flow, ensure reproducibility and support future scalability of this analytical framework.

With these insights and tools, American Airlines is well-positioned to foster innovation, enhance passenger experiences, and achieve sustainable growth in an ever-evolving aviation market.