All Codes and more detailed notes can be found in the python notebook

1. Data Process
   1. Data Preparation (Prepare Data for Processing)
      1. Combine Rows from all Sheets (Combine 5 sheets of data into 1)
      2. Data Cleaning (Cleaning data types and missing values)
         1. Convert Missing Value Type
         2. Drop Missing Rows in Channel Column
         3. Convert Data Format for Modeling
         4. Convert Origin/Dest/OD Columns into Route
         5. Complete Travelling Probability Columns and Create Top Purpose Column
         6. One-Hot Encode Channel and Travelling Purpose
   2. Feature Engineering (Engineer New Features for Modeling)
      1. Date
         1. PNR Create Date
         2. PNR Depart Date
      2. Advance Purchase Date
      3. Round Trip
      4. Distance
      5. Revenue
      6. Travelling Alone
      7. Bin Advance Purchased Time
         1. Advance Purchase Day
         2. Advance Purchase Month
2. Modeling
   1. Data Load
   2. Modeling
      1. Determine the Number of Clusters (WCSS and Elbow Method)
      2. Actual Model (Kmeans with nCluster=4)
      3. Accuracy of Kmeans Model
   3. Summarize Clusters
      1. Averages of Clusters
      2. Number of Observations in Clusters
   4. Graph and Insight about Clusters (Understand the distribution in each cluster)
      1. RPP
         1. Ratio to MedRPP Pies
         2. Ratio to MedRPP Bars
      2. Channel Pies
      3. Round Trip Pie
      4. Travel Alone Pies
      5. Travelling Purpose Pies
      6. Advance Purchase
         1. Advance Purchase Day Stacked Bar Charts
         2. Advance Purchase Month Stacked Bar Charts
   5. Visualization of Clusters