**Competition1**

**DATA VISUALIZATION**

**Name: Anushareddy Ramachandra Reddy**

**Datasets:**

1: Top12\_airlines\_reviews

2: Airline\_top12\_list

**1)Data Integration and Visualization**:

**A screenshot of a computer

Description automatically generated**

The **"Top12\_airlines\_reviews"** and **"Airline\_top12\_list"** datasets that the instructor gave were combined for the analysis.

The Airline\_top12\_list Dataset contains information about the top 12 Airlines and Top12\_airlines\_reviews Dataset contains customers review about airlines.

**A screenshot of a computer

Description automatically generated**

I have used Airlines, URL and country variables in the above visualization to represent the data, Number of reviews is a new calculated field of formula - COUNT([Review Text]) outputs count of reviews.

**Calculation:**

**A screenshot of a computer

Description automatically generated**

**2)Cabin Class Classification**:

**A screenshot of a computer

Description automatically generated**

To distinguish between the different kinds of cabin classes mentioned in reviews, utilize the classification of cabin classes. The following categories such as Business, Economy, First Class, and Unknowns are used for classification. Classify a review as "Unknowns" if it does not state the type of cabin.

**Table Data: Cabin class classification**

Unknowns: No filling in

Business: Traveled in business class

Economy: Traveled in Economy class

First Class: Traveled in First Class

Created a calculated field Cabin Class to distinguish Cabin class type.

**Calculation:**

A screenshot of a computer program

Description automatically generated

**3)Aspect Rating Extraction**:

A screenshot of a computer

Description automatically generated

The task of Aspect Rating Extraction in Natural Language Processing (NLP) entails identifying and obtaining ratings or evaluations linked to different aspects. This is commonly employed in customer feedback analysis, sentiment analysis, and opinion mining. In this exercise, ratings from several viewpoints specified in the Aspect ratings such as legroom, cleanliness, customer service, and check-in and boarding,In-flight Entertainment and Food and Beverage will be identified and derived. Finding and extracting numerical ratings or qualitative judgments pertaining to these features from each review's text will be necessary for this. I've included variables such legroom, customer service, cleanliness, check-in and boarding, In-flight Entertainment and Food and Beverage are aspect ratings in the visualization above. The data is shown in text tables style. Airlines.

**Calculations:**

**Cleanliness:**

A screenshot of a computer

Description automatically generated

**check-in and boarding:**

A screenshot of a computer program

Description automatically generated

**Legroom:**

A screenshot of a computer program

Description automatically generated

**Customer service:**

A screenshot of a computer program

Description automatically generated

**In-flight Entertainment**

A screenshot of a computer program

Description automatically generated

**Food and Beverage**

A screenshot of a computer program

Description automatically generated

**4)Timeline Analysis**:

A screenshot of a computer

Description automatically generated

Visualizing and evaluating data throughout time is the focus of timeline analysis. One useful tool for determining which Cabin Class group regularly receives the lowest total ratings over time is a timeline analysis. This analysis should provide a visual representation of variations and patterns in ratings over different periods of time. The overall ratings Review from 2016 to 2020 are shown in a line graph. The line graph shows that from 2016 Q1 (Quarter 1) to 2019 Q4, every cabin class received an overall review rating of at least 4, with the Unknow Cabin class receiving the highest rating. The review overall rating was highest in 2016 Q3 and 2017 Q1, with a rating of 5, while the rating plummeted between 3 and 2 in 2020 Q1 for all cabin classes, and the lowest in 2019 Q3 for unknow class.

**5) Box-Whisker Analysis**:

The distribution of ratings over a number of time periods together with the traveler's experiences is shown in detail by the Box-Whisker analysis.

A screenshot of a computer

Description automatically generated

This visualization analyzes customer reviews for different airlines over time using a box-whisker plot. The minimum, first quartile (Q1), median, third quartile (Q3), and maximum are the five summary statistics that are used to illustrate the distribution of datasets. It helps with understanding data distribution and identifying outliers. The average review score is displayed on the y axis, and months and quarters are indicated on the x axis. Each airline is depicted using a different hue, and data points show the diversity and central trend of customer happiness. The summary indicates a wide range of scores, with some months exhibiting more review consistency than others. Airlines use this type of study to assess performance over time and in various time periods.