Data Visualisation

CS5803

Individual Courswork Assessment: Tableau Implementation

ID: 2371131 Academic year 2023/24

Table of Contents

Introduction:

With the increase in data the need for analysing the data is increasing day by day, data visualisation is an important technique that helps many business leaders, executives and analysts to help generate valuable insights in the line of their work and helping organisations in creating more value to the customers and investors.

There are millions of travellers travelling across globe every day and everyone has a unique experience when travelling through an airline starting from ticket booking till reaching their destination. Typically Airlines or third parties on behalf of airlines collect feedback from their passengers to measure their experience in areas like ticket booking, boarding, food served on plane etc. This would in turn be use to improve the experience of the other passengers.

In this project, we have used the data visualization techniques learned over the course to create an interactive dashboard on an Airline dataset which contains the user reviews and ratings for different airlines across multiple categories and at an overall experience.

DataSet

The dataset chosen for this project contains the user reviews for the top10 rated airlines in 2023. The reviews cover various aspects of the flight experience, including seat comfort, staff service, food and beverages, inflight entertainment, value for money, and overall rating. This is taken from Kaggle website but the original data source being the Airline Quality website (https://www.airlinequality.com)

The original dataset contains over 8000 records and 17 variables, but as part of the data visualisation project we would not need the text columns related to reviews and so they have been removed.

Column Name	**Column Description**	**Data Type**
Name	Name of the passenger who provided the review	String
Review Date	The date on which review is given	Date
Airline	The Airline name - Singapore Airlines, Qatar Airways, ANA (All Nippon Airways), Emirates, Japan Airlines, Turkish Airlines, Air France, Cathay Pacific Airways, EVA Air, and Korean Air	String
Verified	If the review is coming from a verified user or not. - True/False	Binary
Type of Traveller	Type of traveller - Solo/Couple/Family/Business	Nominal
Month Flown	Month of Travel	String
Route	Travel route, source and destination	Nominal

Class	Seat class - Business/ Economy/ First/ Premium Economy	Nominal
Seat Comfort	Rating for seat comfort out of 5	Ordinal
Staff Service	Rating for Staff Service out of 5	Ordinal
Food & Beverages	Rating for Food and beverages out of 5	Ordinal
Inflight Entertainment	Rating for Inflight Entertainment out of 5	Ordinal
Value For Money	Rating for Value for Money out of 5	Ordinal
Overall Rating	Overall rating of the experience with the Airline given out of 10	Ordinal
Recommended	Whether passenger recommends the given airline to others or not - Yes/No	Binary

Data Source	https://www.kaggle.com/datasets/sujalsuthar/airlines-reviews (Original datasource - https://www.airlinequality.com)
Author	Sujal Suthar
Last updated	February 2024

Persona:

For the purpose of this project we would like to take the Airline Customer Success manager of Singapore Airlines as intended persona, who would be interested in looking at the ratings for a the Airline and how it is doing as compared to other Airlines.

Persona: Customer Success Manager - Singapore Airlines.

Questions Formulation and User requirements

Questions:

Q1. How is the overall rating of the Singapore Airlines as compared to the other Airlines?
Q2. How is overall rating changed over the years for Singapore Airlines.
Q3. How is Singapore airlines faring on overall rating across classes over the given period?
Q4. How is the distribution of rating for different aspects of flight experience vary for the verified users across different types of travellers for Singapore Airlines.
Q5. Analyse how the different aspects of the flight experience are varying for those who has rated the Singapore airlines as one/two overall.

User requirements:

R1. To answer the Q1, the user needs to visualise variation of the average overall rating for each of the airlines. A good view would be the bar chart showing the average overall rating on the y-axis and the Airlines on the x-axis. With the Average overall rating shown as label on top of each bar.

R2. To answer Q2, user needs to visualise the trend line showing the average overall rating variation over the given period. For this a time-series line graph is needed, with average overall rating shown on y-axis and the year from the review date on the x-axis.

R3. To answer Q3, user needs to visualise the trendlines of average overall rating variation across classes over the given period. For this a time-series line graph is needed, with overall rating is shown on the y-axis and the year from the review date on the x-axis. The class variable is used in colour marks to differentiate the class wise trends.

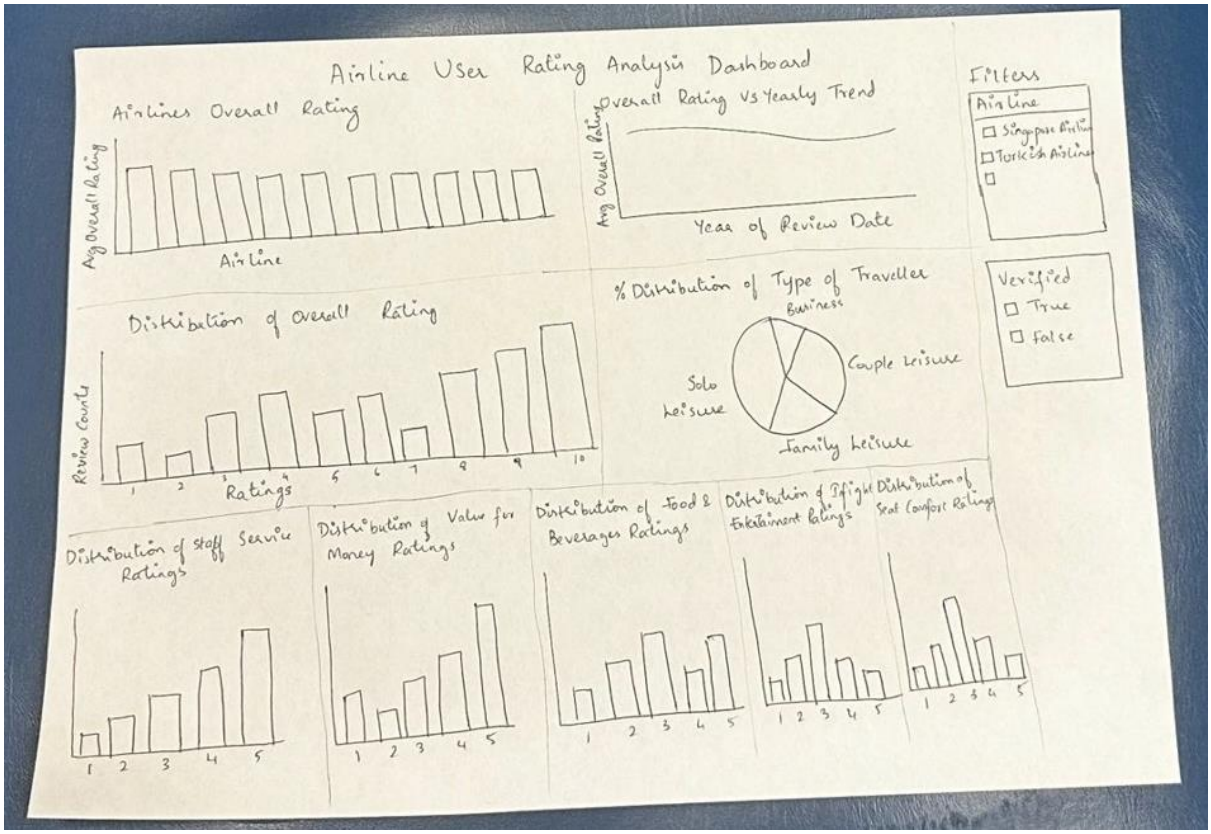
R4. To answer Q4, user needs to visualise the distribution of the user counts at the rating level which is rating levels 1, 2, 3, 4,5. Across different flight experience aspects like Seat Comfort, Staff Service, Inflight Entertainment, Food and beverages and Value for money for the verified users while checking for each type of traveller. A good view of this is first to create the pie charts for each of the Flight experience aspects separately with user ratings on the x-axis and counts on the y-axis.
Similarly we need to check the distributions against the overall rating levels and should have the option to view and give filters based on the overall rating levels 1 to 10. So a bar graph is drawn for overall rating as well.
Second give filters options for verified user and type of traveller selection. Instead of creating type of traveller as filter we could also draft a pie chart for type of traveller as this gives the option to help filter the user from the dashboard when player clicks on any of the given category for type of traveller.

Design:

To answer the above questions, overall we would have three types of charts – Bar chart, time series line graph and pie chart.

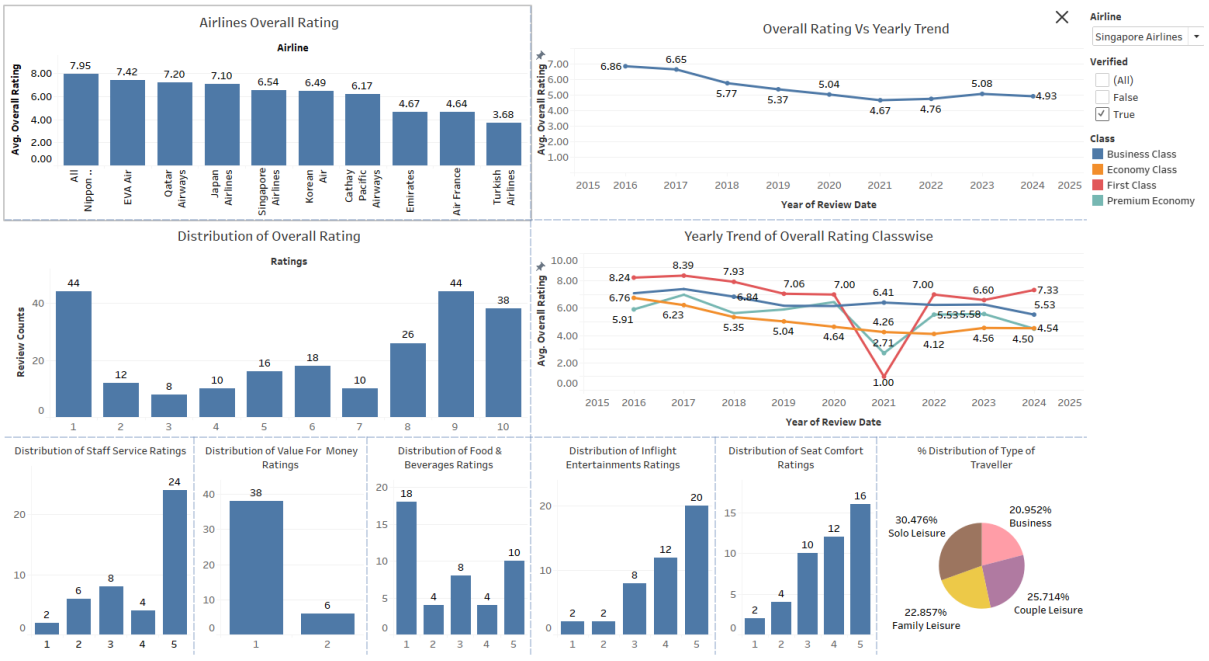
Initially only Q1, Q2 and Q4 were part of the scope so in total we had 6 barcharts, one time series chart and one pie chart.

A dashboard prototype was built in order to make sure the design is coming out correctly and it is user friendly. Below is the prototype which was initially designed.



As we were progressing we have included Q3 as we thought that it might be a relevant chart for the customer success manager to understand the trends of ratings at the class level. As tableau supports the colour encoding for the variables with a graph, colour encoding can be used to differentiate the classes in the time series graph.

The final dashboard has evolved as below.

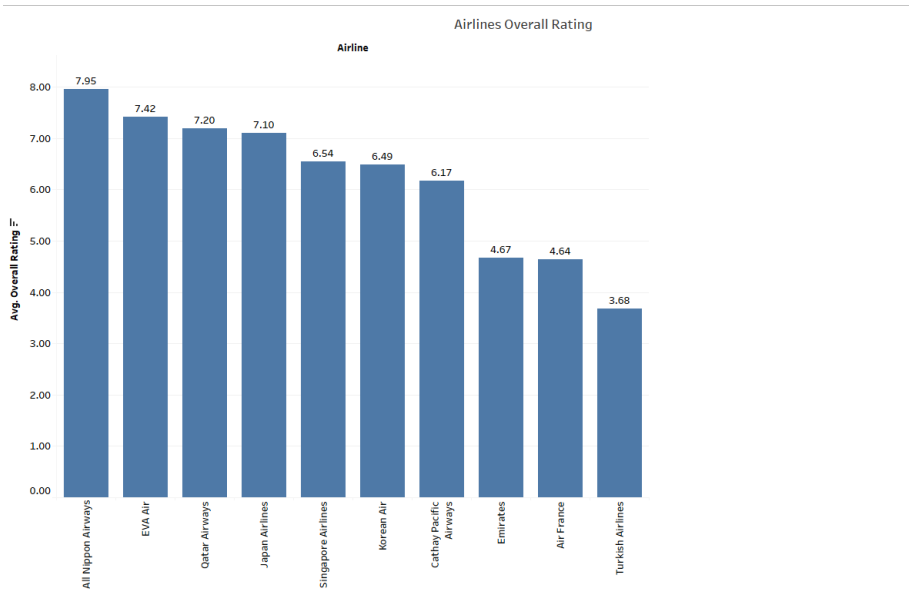


Implmentation

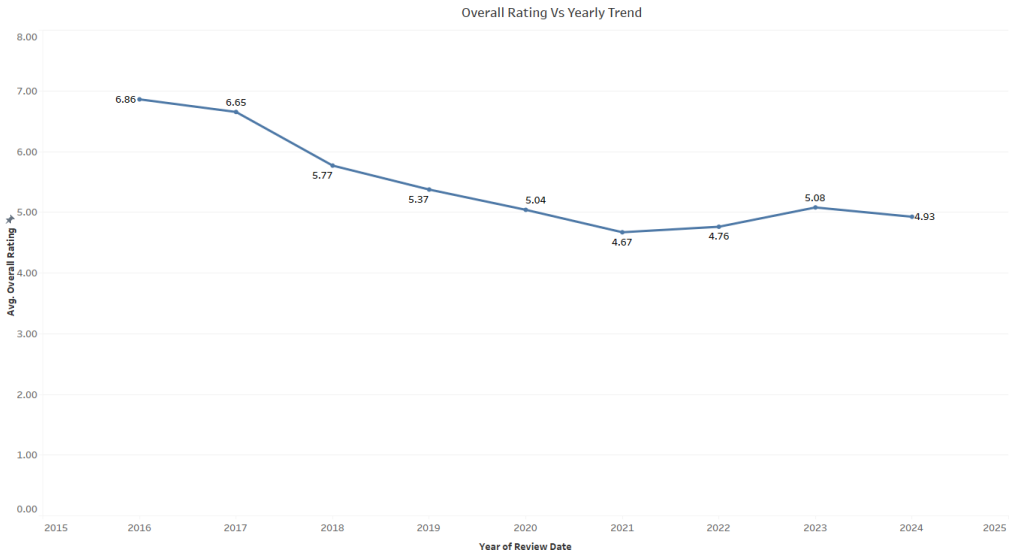
Data loading: Data connection is established to the csv file from the tableau and the data is loaded. Tableau detects the data types and categorises them as measures and dimensions.

In our case even though the variables Overall rating is ordinal in nature it can be treated as continuous to calculate the average overall rating which is a measure to know how the users perceive the airline, this is a standard process to consider the average of the ratings in any industry.

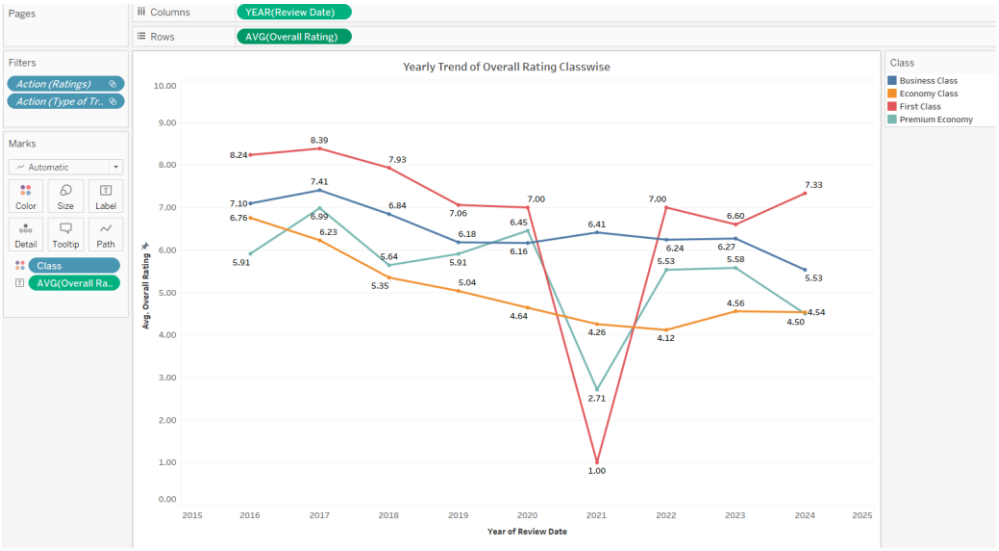
Q1. To implement the bar chart with Airline as x-axis and average overall rating as y-axis. As Airline is a categorical variable as per bertins law we need to consider the bar-chart as it's the best practise for visual encoding.
Steps involves first dragging the Airline into column tab and Overall rating into rows column. By default tableau aggregates the values by considering the SUM, but this has to be changed to see the Average. The diagram looks as below.



Q2. To implement the time series trend, we need to drag the review date on to the columns tab and choose the Year level from the dropdown on the Review date parameter.
The overall rating is dragged to rows and the aggregation is changed to Average.
Next the Overall rating is given as label to see the values over the line.



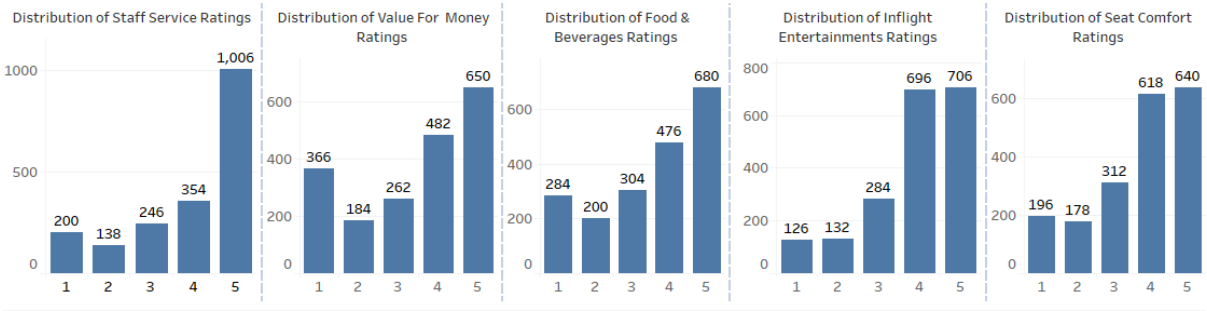
Q3. This one is implemented same as Q2 but that class variable is used as legend so that the individual class representation is see in the same view to understand how each class is faring compared to other over the time period.

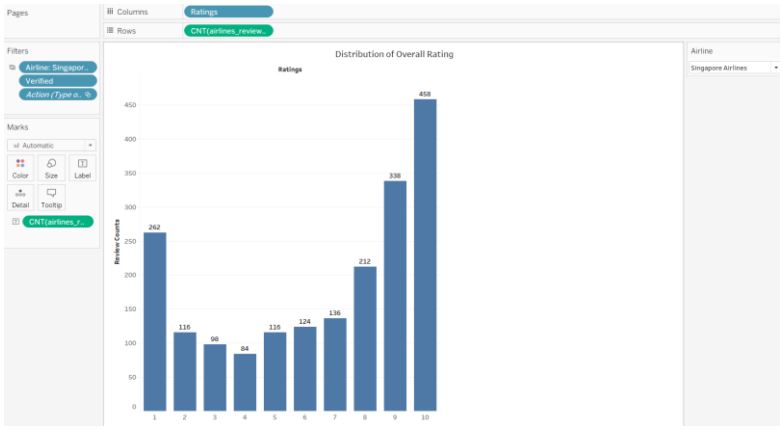


Q4. To implement this first we have to make sure the columns Seat comfort, Inflight Mangement, Staff Service, Value for Money and Food&Beverages are converted to dimensions so the rating values are treated as categorical. We now have create 4 bar charts one each for each variable.
For example for Staff service barchart drag the staff service variable into the columns and then drag the records count variable into the rows.
Repeat this for all the other 4 variables of the flight experience columns Seat comfort, Inflight Mangement, Value for Money and Food&Beverages respectively and also for the overall rating to see the similar graph.

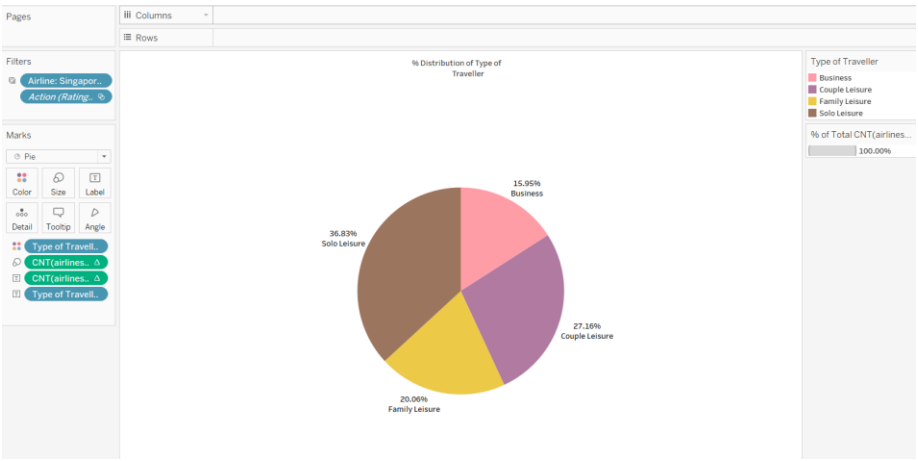
The most important factor is how we make use of the filter utilizations. The filters have to applied at multiple work sheets level so in the dashboard we have to change the setting to apply

one filter across multiple sheets in this case the verified filter has to be applied for all the other sheets.





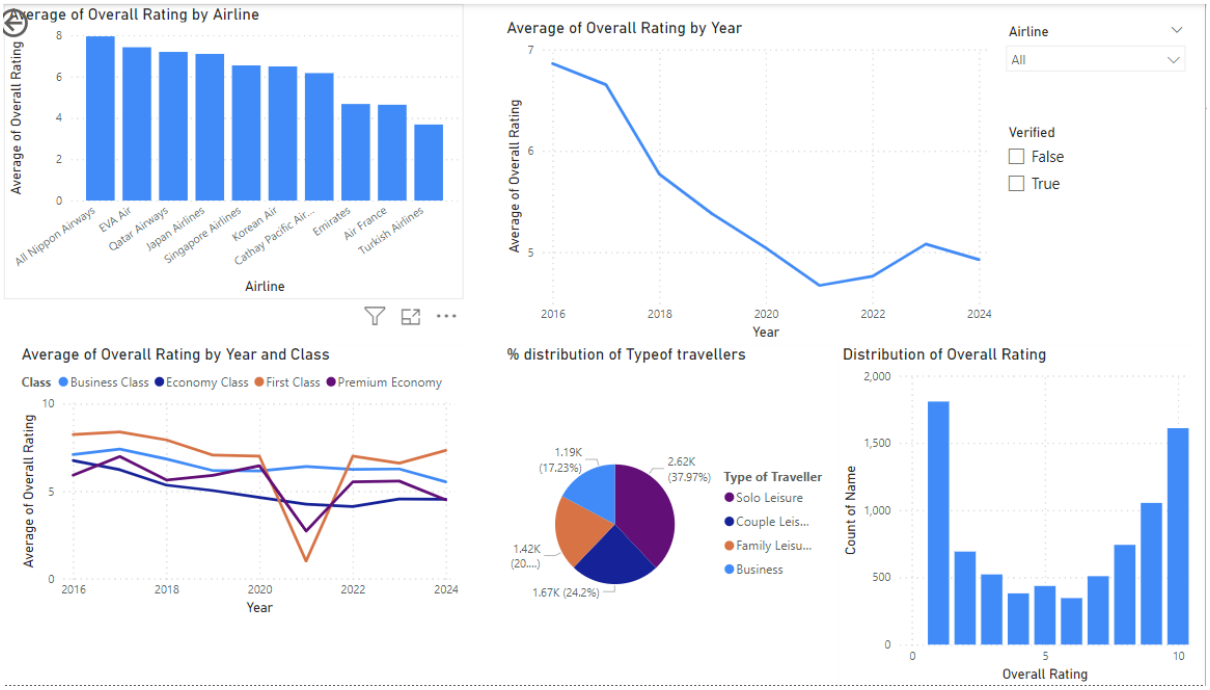
Next graph would be the pie chart that will be created using the 'Type of traveller' and the record counts variable. The pie chart would be selected and both the variables are dragged into the Marks tab. Now the record counts have to changed to calculated field by using the quick calculations and changed to 'Percent of total'. To display the percentage distribution of the 'Type of Travellers'.



We would want to use this pie chart as a filter as well to answer the question so this has to be applied as filter for all the relevant sheets in the dashboard worksheet.

PowerBI implementation:

The powerBI implemention is done for few of the charts but not for the full. It is observed that the PowerBI tool has a different approach to the way the visualisations are done. The data transformation options are good for powerBI but the automatic detection of data types and conversion is not that straight forward as tableau. Below views are attempted it PowerBI to answer the questions Q1, Q2, Q3 and partially Q4



Walk Through:

Q1.
A1. From the visualization it can be see that Singapore airlines stand at 5th position in the average overall ratings. The average overall rating is almost 1.5 points less than the top. It is better than the Airlines Korean, Cathay Pacific, AirFrance and Turkey

Q2.
A2. The Singapore airline rating is decreasing aver the years from 2015 to 2024, lowest being in 2021. There was an increase from 2021 to 2023. But not at a significant level.

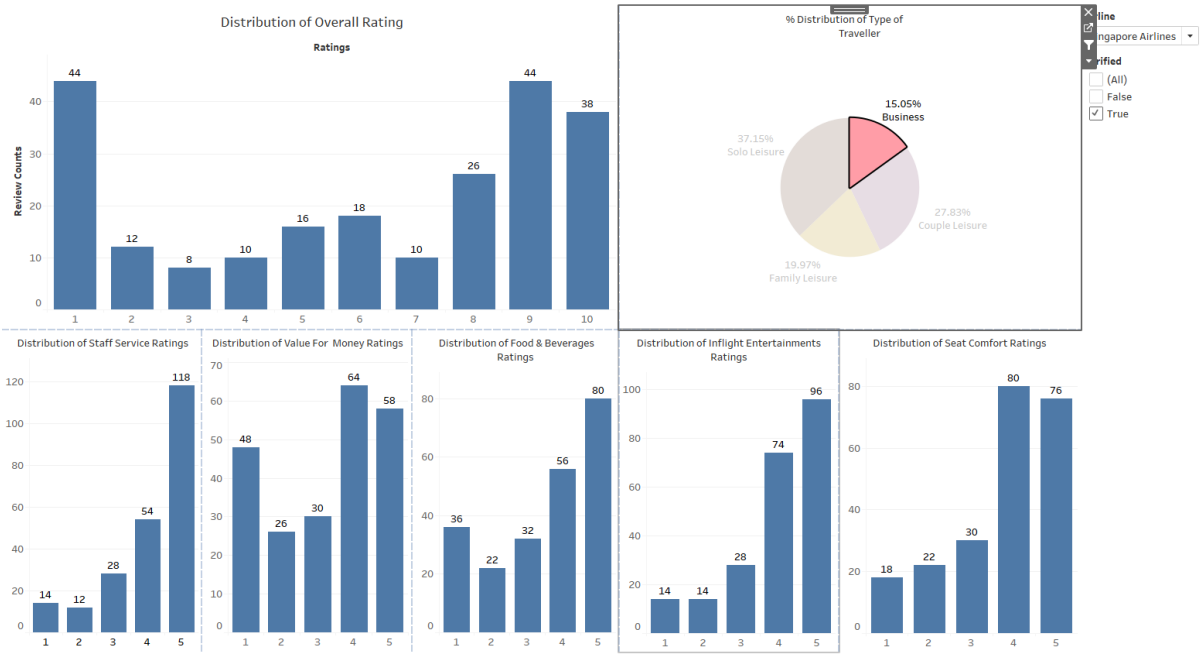
Q3.
A4. Looking at the class wise overall rating trend it can be seen all the classes has a decreasing trend at a high level. But it can be seen that the premium economy and first classes have highest dips in 2021, especially the first class.

Q4

A4. It can be seen when from the below business type verified users have given overall bad ratings, the number of people who gave rating 1 is greater than users who gave rating 10 that means business travellers have very bad experience.

On a further drilling looking at the distribution of ratings on these 44 users it can be see that 38 users have rated 1 on value for money and 18 rated 1 on Foods and Beverages.

This means Business travellers are highly dissatisfied in 'Value for Money' and 'Food and Beverages'. These are high value customers who cannot be ignored, the customer service has to be really really improved by Singapore airlines. So the customer success manager has to really work to make sure the business travellers are satisfied.



Conclusion and Discussion:

This project has helped overall to understand the data visualization techniques that needs to be adopted for data analysis. The concepts learnt with respect to the encoding, representation and presentation are used where applicable and a tableau dashboard is created on the Airline user ratings data set. A dashboard is created keeping in view of the persona Customer success manager of the Singapore airlines to help access him understand where does the airline stands interms of the ratings and what all can be improved given the data. One good outcome is that the business class verified users are highly dissatisfied with airlines value for money and Food and beverages offered to them. During the implementation of this project it has been a great learning as a course student for me on how to effectively design good dashboards in tableau, particularly the learning of effective use of filters across the dashboard in an efficient way, story writing and also effective space utilization. These learned concepts can be applied to effectively solve real world problems which can be solved by data visualization.

Appendix:

The appendix material is submitted as a zipped/archived folder containing the following files:

1. airline_reviews.csv
2. CS5803_2371131_Tableau.twb
3. CS5803_2371131_PowerBI.pbix