Project Objective

A US-based airline headquartered in Boston, Massachusetts has just sent the latest passenger survey and it looks like the **satisfaction rate dipped under 50%** for the first time ever. The leadership team needs to take action fast, so they've brought you in to analyze the data and find the key areas to focus on for getting back on track.

Your task is to **recommend a data-driven strategy for increasing the Airlines' satisfaction rate**, and present it in the form of a single page report or dashboard.

About The Data Set

Customer satisfaction scores from 120,000+ airline passengers, including additional information about each passenger, their flight, and type of travel, as well as their evaluation of different factors like cleanliness, comfort, service, and overall experience.

The data contained various personal information on each passenger, as well as details about their journeys and class of travel

In addition, there were a number of questions which ask to rate particular services on a scale of 1 to 5, with 1 being the worst service and 5 the best (see below from the data dictionary).

Field	Description
ID	Unique passenger identifier
Gender	Gender of the passenger (Female/Male)
Age	Age of the passenger
Customer Type	Type of airline customer (First-time/Returning)
Type of Travel	Purpose of the flight (Business/Personal)
Class	Travel class in the airplane for the passenger seat
Flight Distance	Flight distance in miles
Departure Delay	Flight departure delay in minutes
Arrival Delay	Flight arrival delay in minutes
Departure and Arrival Time Convenience	Satisfaction level with the convenience of the flight departure and arrival times from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Ease of Online Booking	Satisfaction level with the online booking experience from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Check-in Service	Satisfaction level with the check-in service from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Online Boarding	Satisfaction level with the online boarding experience from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Gate Location	Satisfaction level with the gate location in the airport from 1 (lowest) to 5 (highest) - 0 means "not applicable"
On-board Service	Satisfaction level with the on-boarding service in the airport from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Seat Comfort	Satisfaction level with the comfort of the airplane seat from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Leg Room Service	Satisfaction level with the leg room of the airplane seat from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Cleanliness	Satisfaction level with the cleanliness of the airplane from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Food and Drink	Satisfaction level with the food and drinks on the airplane from 1 (lowest) to 5 (highest) - 0 means "not applicable"
In-flight Service	Satisfaction level with the in-flight service from 1 (lowest) to 5 (highest) - 0 means "not applicable"
In-flight Wifi Service	Satisfaction level with the in-flight Wifi service from 1 (lowest) to 5 (highest) - 0 means "not applicable"
In-flight Entertainment	Satisfaction level with the in-flight entertainment from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Baggage Handling	Satisfaction level with the baggage handling from the airline from 1 (lowest) to 5 (highest) - 0 means "not applicable"
Satisfaction	Overall satisfaction level with the airline (Satisfied/Neutral or unsatisfied)

This type of survey is typically known as a Likert Scale survey question. Here a 5 number scale has been used, but you may also encounter 7 or 10 number scales. You may also come across non-numeric versions (e.g. “very likely” to “not likely at all”).

These types of scales allow for a little more nuance in the sentiment surrounding a survey response where the answer may not be a binary yes/no.

Analysis of Likert Scales

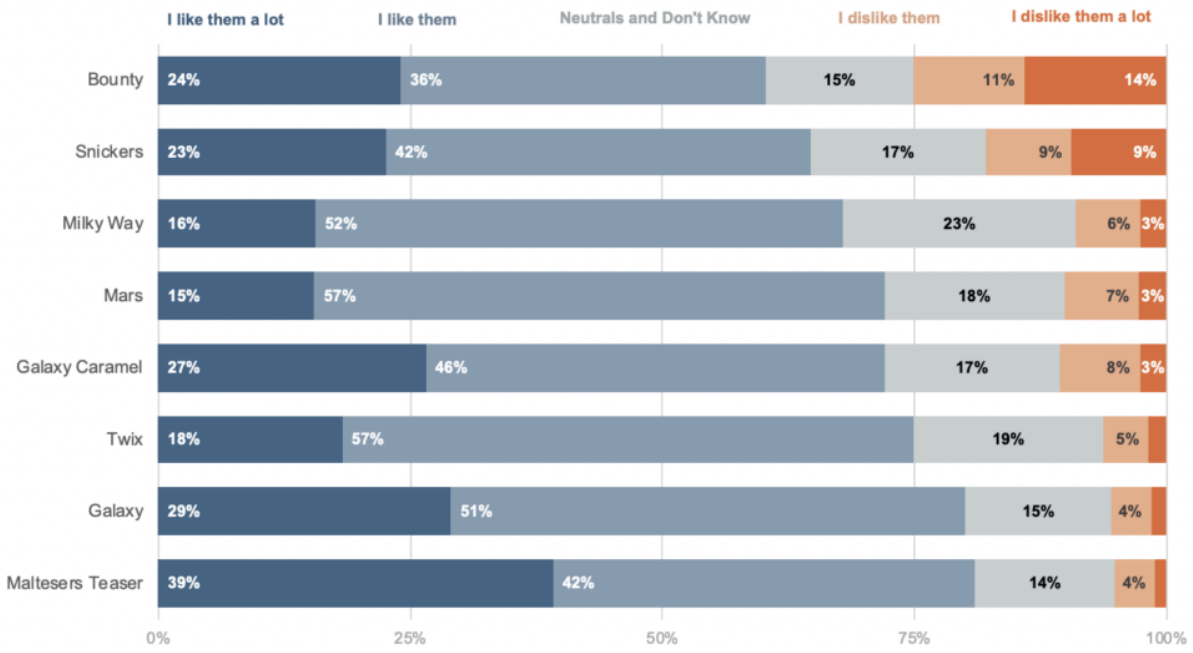
I will not deep dive into the theory of Likert Scales, but as I have endlessly researched, tried to understand this concept and from theory convert into a practical analysis, I came across many articles, Youtube to see how they handle the Likert Analysis.

Below are a few different ways of presenting data using the scales. This ranges from a simple stacked bar chart, to basing it around a neutral zero value, to extracting neutral values, and all the way to a full deconstruction.

Each method has its own particular advantages depending on what you are trying to present, but they are effective at showing the relative distribution of a spectrum of responses.

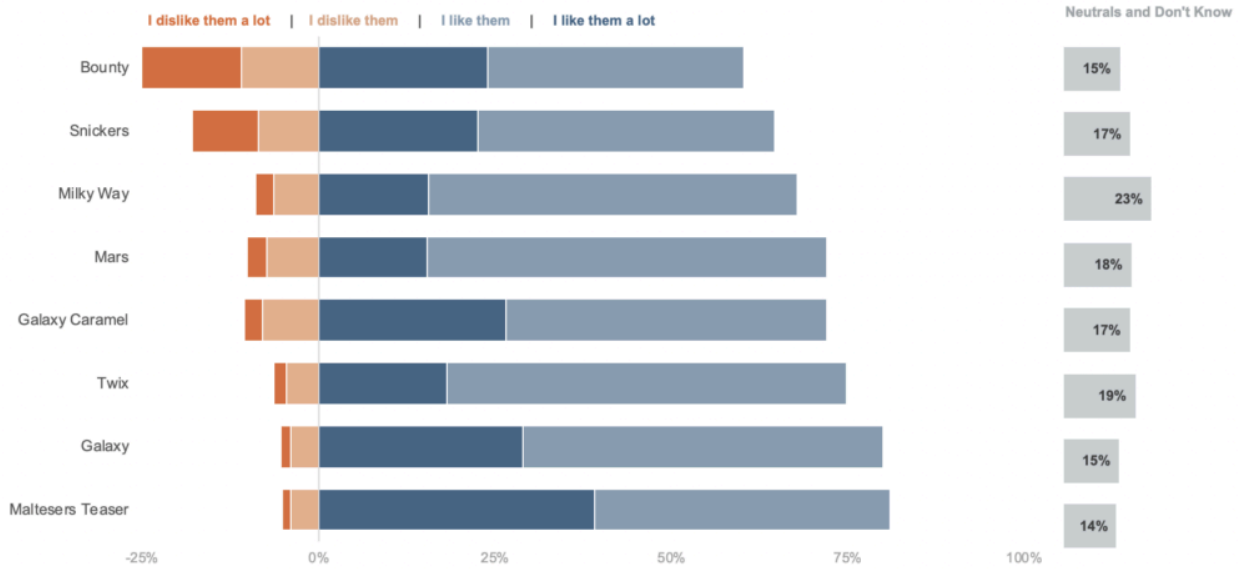
Everyone likes chocolates, but Bounty and Snickers get the most extreme opinions

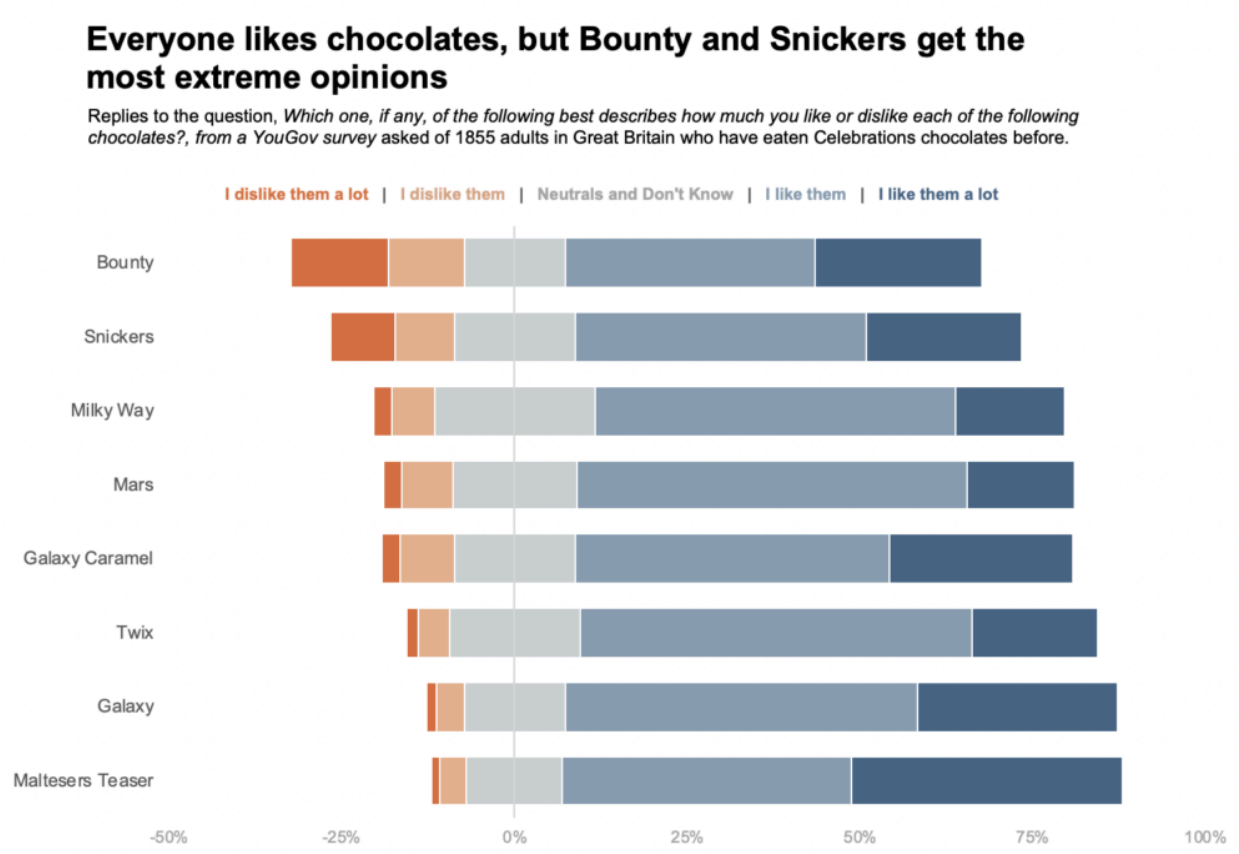
Replies to the question, *Which one, if any, of the following best describes how much you like or dislike each of the following chocolates?, from a YouGov survey* asked of 1855 adults in Great Britain who have eaten Celebrations chocolates before.

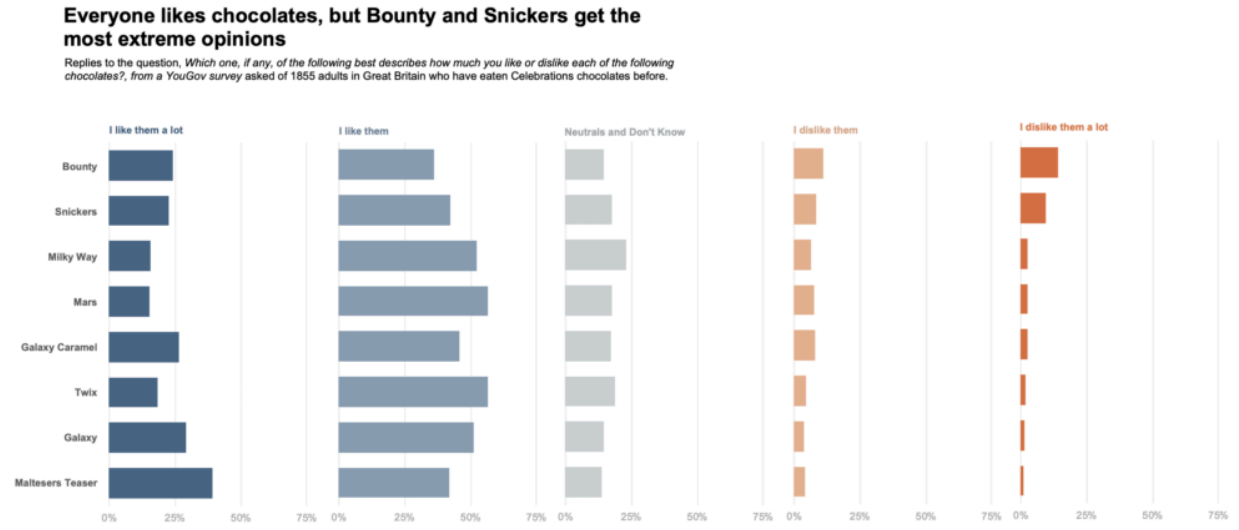


Everyone likes chocolates, but Bounty and Snickers get the most extreme opinions

Replies to the question, *Which one, if any, of the following best describes how much you like or dislike each of the following chocolates?, from a YouGov survey* asked of 1855 adults in Great Britain who have eaten Celebrations chocolates before.

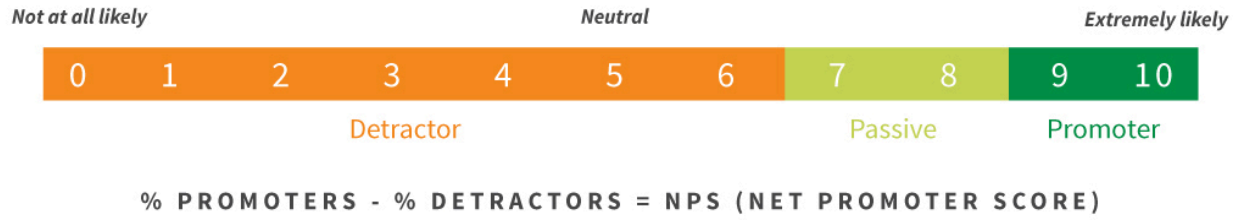






Net Promoter Score

A child of the Likert scale is the **Net Promoter Score (NPS).** Strictly speaking, this is used on survey results with scales of 1-10, and it is calculated by subtracting the % of promoter scores (9-10) from the % of the detractor scores (0-6)



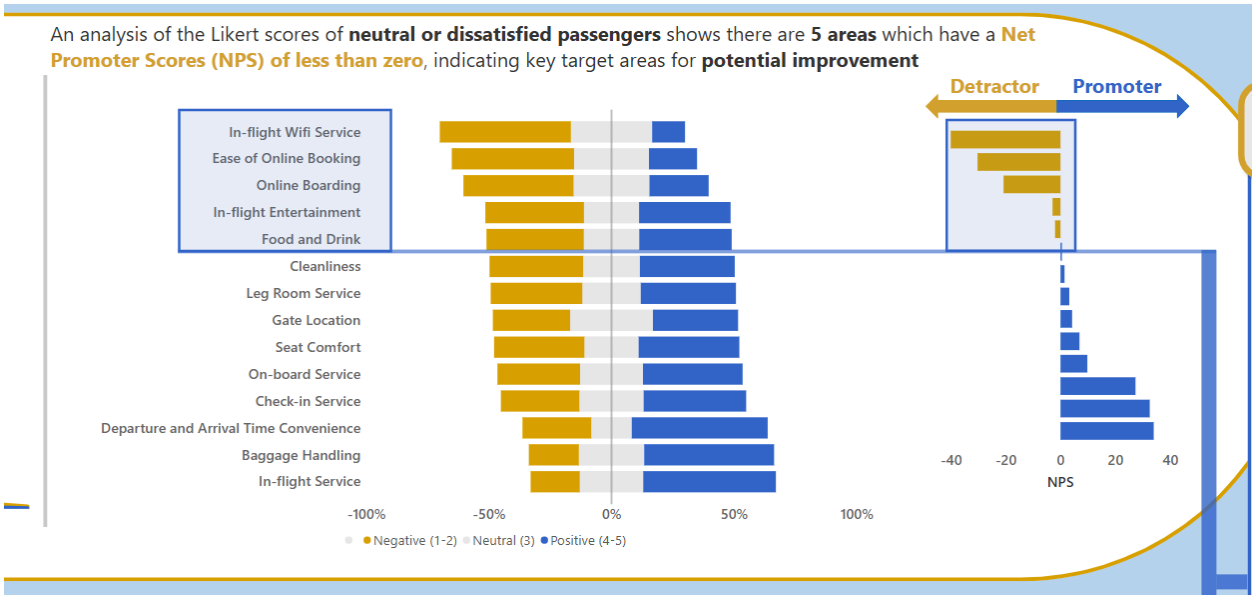
% PROMOTERS - % DETRACTORS = NPS (NET PROMOTER SCORE)

If you return a positive result, then people are more enthusiastic about that aspect of your brand or company, whereas if it is negative, that could be damaging to your company

What I wanted to show

Taking this into account, here is the analysis version that I want to show. However, contradictory to the theory mentioned above, the dataset scale is only from 1 to 5, so I need to rescale the analysis based on the following:

- 1-2 = Negative
- 3 = Neutral
- 4-5 = Positive

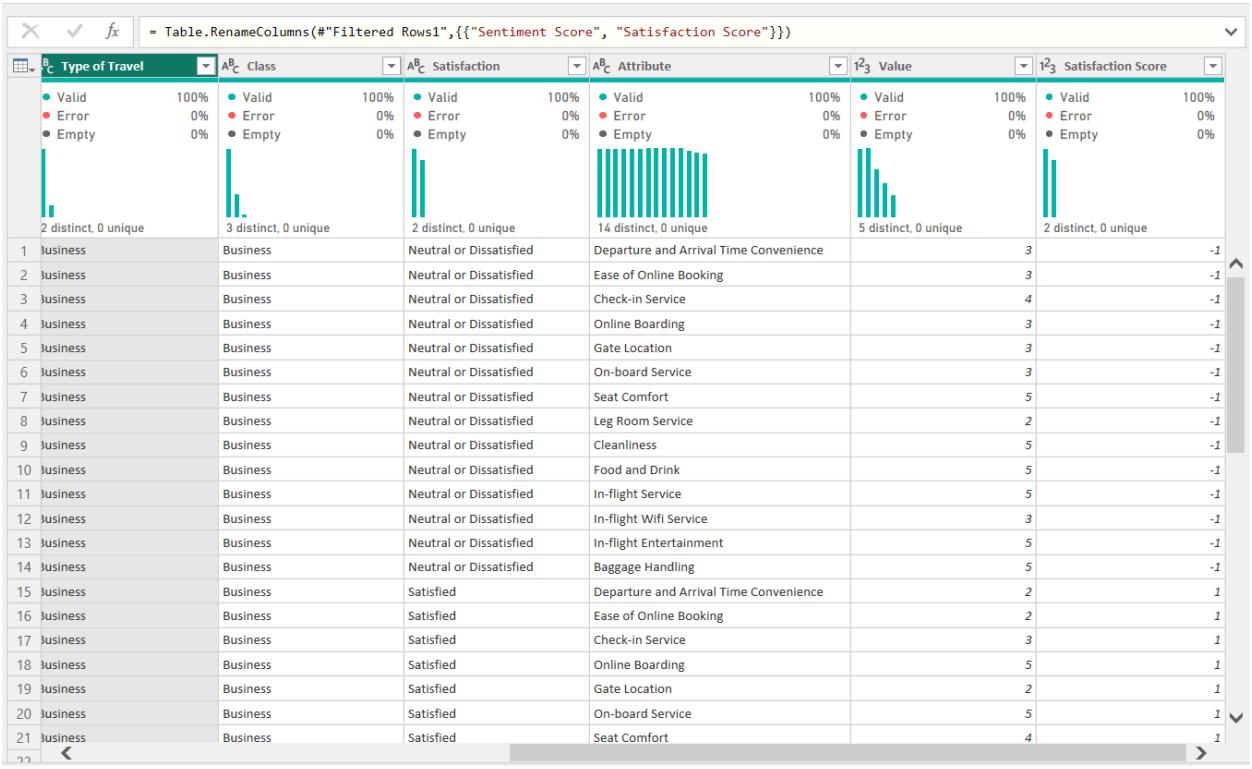


How to Handle it on Power BI

1. Power Query Work

First I imported the dataset into Power Query, then created a duplicate table of the original table. Next, I selected the column headers for all the columns containing the Likert categories. After

that, I navigated to the Transform ribbon at the top, then selected “Unpivot Columns” → “Unpivot Selected Columns”



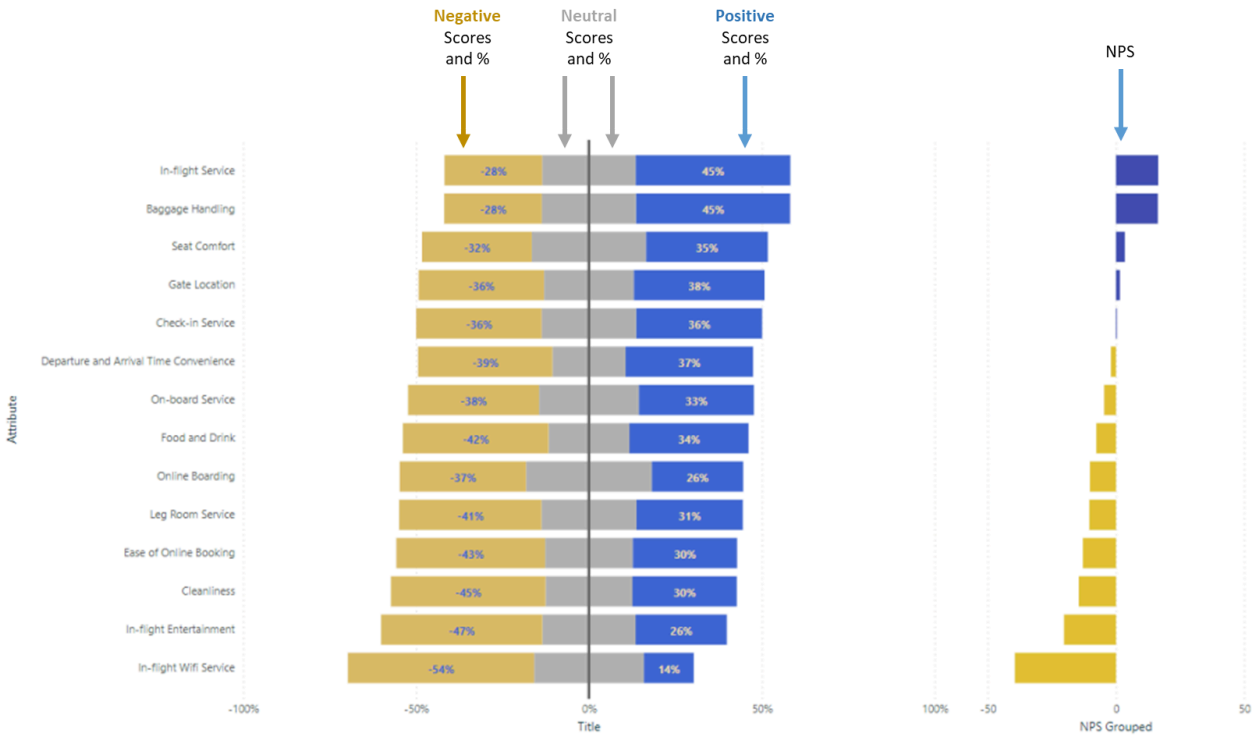
	Type of Travel	Class	Satisfaction	Attribute	Value	Satisfaction Score
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
	2 distinct, 0 unique	3 distinct, 0 unique	2 distinct, 0 unique	14 distinct, 0 unique	5 distinct, 0 unique	2 distinct, 0 unique
1	Business	Business	Neutral or Dissatisfied	Departure and Arrival Time Convenience	3	-1
2	Business	Business	Neutral or Dissatisfied	Ease of Online Booking	3	-1
3	Business	Business	Neutral or Dissatisfied	Check-in Service	4	-1
4	Business	Business	Neutral or Dissatisfied	Online Boarding	3	-1
5	Business	Business	Neutral or Dissatisfied	Gate Location	3	-1
6	Business	Business	Neutral or Dissatisfied	On-board Service	3	-1
7	Business	Business	Neutral or Dissatisfied	Seat Comfort	5	-1
8	Business	Business	Neutral or Dissatisfied	Leg Room Service	2	-1
9	Business	Business	Neutral or Dissatisfied	Cleanliness	5	-1
10	Business	Business	Neutral or Dissatisfied	Food and Drink	5	-1
11	Business	Business	Neutral or Dissatisfied	In-flight Service	5	-1
12	Business	Business	Neutral or Dissatisfied	In-flight Wifi Service	3	-1
13	Business	Business	Neutral or Dissatisfied	In-flight Entertainment	5	-1
14	Business	Business	Neutral or Dissatisfied	Baggage Handling	5	-1
15	Business	Business	Satisfied	Departure and Arrival Time Convenience	2	1
16	Business	Business	Satisfied	Ease of Online Booking	2	1
17	Business	Business	Satisfied	Check-in Service	3	1
18	Business	Business	Satisfied	Online Boarding	5	1
19	Business	Business	Satisfied	Gate Location	2	1
20	Business	Business	Satisfied	On-board Service	5	1
21	Business	Business	Satisfied	Seat Comfort	4	1

This resulted in the above layout. There is a column called “Attributes”, which relates to all the Likert categories columns/questions, and another column called “Values” which are the corresponding Likert scale numbers - effectively all the categories and their corresponding values are in two columns

2. Measures

Now that the organization of the data was done in Power Query, I moved on to start creating the measures that would help me build the below draft visual. This include the:

- Count the positive scores (4-5)
- Count the negative scores (1-2)
- Count the neutral scores (3) - this was done in two parts as they straddled the zero line
- % positive
- % negative
- % neutral
- NPS



These DAX measure are followed as below:
COUNT

```
Count Scores =
COUNTA(
    Pass_Satisfaction_Scores[Value]
)
```

```
Count Score - Positive =
(CALCULATE(
    [Count Scores],
    FILTER(Pass_Satisfaction_Scores, Pass_Satisfaction_Scores[Value] IN {4,5}
    )
))
```

Note that the below calculation is multiplied by -1 enable negative side of the y-axis

```
Count Scores - Negative =
-1 * (CALCULATE(
    [Count Scores],
    FILTER(Pass_Satisfaction_Scores, Pass_Satisfaction_Scores[Value] IN {1,2})))
```

Next, as mentioned for the neutral scores, these straddle the zero line, therefore the calculation is split into two, one for the positive side, and one for the negative side.

```
Count Scores - Neutral P =
0.5 * (CALCULATE(
    [Count Scores],
    FILTER(Pass_Satisfaction_Scores, Pass_Satisfaction_Scores[Value] = 3)))
```

```
Count Scores - Neutral N =
-0.5 * (CALCULATE(
    [Count Scores],
    FILTER(Pass_Satisfaction_Scores, Pass_Satisfaction_Scores[Value] = 3)))
```

Finally, in order to transform these count measures into % values, I need a denominator. Typically, you could use an ALL() DAX function so that you are measure a portion of the total of a given column, but as I have created an unpivoted table with multiple categories and slice/dice by flight class (Economy/Economy Plus and Business), it requires something a little more refined.

Instead of using ALL(), I opted for ALLEXCEPT() with several columns referenced:

```
All Selected Attribute =
CALCULATE(
    [Count Scores],
    ALLEXCEPT(Pass_Satisfaction_Scores, Pass_Satisfaction_Scores[Attribute], Pass_Satisfaction_Scores[Class],
    Pass_Satisfaction_Scores[Satisfaction]
    )
)
```

Percentage Values

Now this was done, all that was left to do was to create a few simple **DIVIDE()** functions to finalize the percentage values - simply dividing each count by the **All Selected Attribute**

```
% Scores Positive =
DIVIDE(
    [Count Score - Positive], [All Selected Attribute]
)
```

```
% Scores Negative =
DIVIDE(
    [Count Scores - Negative], [All Selected Attribute]
)
```

```
% Scores 3P =
DIVIDE(
    [Count Scores - Neutral P], [All Selected Attribute]
)
```

```
% Scores 3N =
DIVIDE(
    [Count Scores - Neutral N], [All Selected Attribute]
)
```

For NPS calculation, after all those measures were created, I can now come to calculating the NPS. If we relook at the below calculation, I recreated a version using my measures

```
NPS Grouped =
100*(([% Scores Positive]+[% Scores Negative]))
```

Then in order to order my Likert chart by the NPS. I used the **RANKX** function to rank the attributes by the NPS in descending order

```
Rank NPS =
RANKX(
    ALL(
        Pass_Satisfaction_Scores[Attribute]
    ),
    [NPS Grouped],,
    DESC
)

```