Northumbria University
NEWCASTLE
Newcastle • London • Amsterdam

# COURSEWORK COVER PAGE

| | |
|---|---|
| **Module Number:** | LD7186 |
| **Module Title:** | Big Data Analytics |
| **Module Tutor Name:** | SYED RAZA [SR] |
| **Coursework Title:** | Assignment |
| **Student Name:** | RASIKH SADIQ THAKUR |
| **Student ID:** | W20342021 |
| **Programme of Study:** | MSc Big Data & Data Science Tech [BD] |
| **Word count:** | 3626 |

**Submission Instructions**

1. Name your submission in according to the name convention, LD7187_<your tutor initial>_<your programme><sem>_<your id><your first name>.docx, eg LD7187_RBS_CT2_w22012345John.docx is the filename for the student enrolled in MSc Computing & Technology sem 2 (2023 Sep Intakes) attending Rejwan's session.
2. Submit to Final Report Submission Point at Bb before 16:00, 22 May 2024

**Declaration**

*I confirm that this assessment is my own work and that I have duly acknowledged and correctly referenced the work of others. I am aware of and understand that any breaches to the Code of Academic Conduct will be investigated and sanctioned in accordance with the Academic Conduct Regulation.*

| | | | |
|---|---|---|---|
| **Your signature:** | RASIKH SADIQ THAKUR | **Date:** | 22/5/2024 |

**Table Of Contents**

**Section 1 - Big Data Analytics (Python)**

**Task 1: Problem Domain, Data Description, and Research Questions**

**Problem Domain and Data Description:**
This data set, "Airline_Review.csv", seems to be dealing with the domain of air passengers, review and comments. Well, it is one of the forms of customers relations with airlines since it can collect users' data on the airline type, flight routes, feedback, reviews, and customer experiences.

**Detailed description of the dataset:**
The dataset appears to contain the following columns:

**Passenger_Name:** Name of the passenger.
**Flying_month:** Month of the flight.
**Route:** Route of the flight (departure and arrival destinations).
**Rating:** Rating provided by the passenger (likely on a scale of 1 to 10).
**Verified:** Verification status of the review.
**Review_title:** Title of the review.
**Review_content:** Detailed content of the review.
**Traveller_type:** Type of the traveller.
**Class:** Class of travel (e.g. Economy, Business, Premium Economy).

**Research Questions:**
Based on the dataset, we can formulate the following research questions:

1. What in planning a holiday airline disruption potentiality such as people influences are?
2. What concern should be given to the time of booking which will affect the likes and behavior of customers?
3. Under such circumstances, can we prove that a person buying a ticket will make travel arrangements?
4. The issue of flight timetables and routes that are the most commonly searched by the passengers is also worth considering.

**Null and Alternate Hypothesis:**
**Null Hypothesis(H0):** Customer characteristics like a customer being vacation-class or a tourist type are not the basis for a better vacation reservation.

**Alternate Hypothesis(H1):** Customers might give up the idea of a successful vacation due to the inherent weaknesses of the customer on whom the holiday reservation depends.

**Task 2: Solution Exploration**

**Evaluation of Approaches and Technologies:**

1. **Python with pandas, Numpy, and Matplotlib/Seaborn:**

   Data analysts tend to prefer using Python due to its simplicity + ease of use + and the fact that it has libraries for pandas data manipulation + NumPy for numerical calculations + as well as Matplotlib/Seaborn for data visualization.
2. **Big Data Frameworks like Apache Spark:**
   Big data frameworks are very helpful as they guide us to manage huge data in an effective way. Apachi Spark, for instance, is very good for the fast and profile data processing due to its distribut source programming capabilities.
3. **Machine Learning Algorithms:**
   For predictive modeling tasks, such as computing the probability of successful perspectives using consumer attributes, ML methods can be applied.

**Solutions and Techniques for Similar Problems:**
Diverse approaches have been employed to tackle comparable issues within the realm of consumer feedback research:

1. **Sentiment Analysis:** Based on the analysis of customer reviews and positive customer experience or negative reviews, NLP techniques like sentiment analysis can help in the extraction of information.
2. **Classification Algorithms:** Embracing historical information, logistic regression, and random forests are just some examples of educational machine learning algorithms that utilize such data in the classifications of customer satisfaction or the potential of a booking success.

3. **Topic Modeling:** Tools like topic modelling (for example Latent Dirichlet Allocation) are very useful to find topics or subjects that arise repeatedly in consumer feedback. This gives the airlines a guide where they might need to be redesigned.

**Chosen Methodological Approach:**
Python will be our main tool along with the Pandas, NumPy and Matplotlib/Seaborn libraries do some data processing. This strategy has many benefits:

1. **Flexibility:** Python has this meta property since it is flexible and can be used in the areas of data analytics, visualization, and manipulation, thereby customizing our study to the research questions and dataset.
2. **Rich Ecosystem:** The existence of a large number of Python modules and tools for numerous data analytics tasks affirms that the ecosystem has everything we need to solve the problem.
3. **Ease of Interpretation:** The Python in the teamwork as well as the present results communication is OK, it is less difficult to read and understand.

**Justification:**
The macrotools of Python, which enable the versatility of the language, and the strong library support available in data science and analytics space have made it the leading language. Regardless of the problems we may be faced with in this study, Python and its libraries allow us to dive into the data and preprocess it, then run statistics on it, and finally visualize the outcome.

**Task 3: Solution Development**
First we have imported the necessary libraries and read the Airline_Review.csv data.

```
# Importing necessary packages
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("Airline_Review.csv")
```

df

| | Passanger_Name | Flying_month | Route | Rating | Verified | Review_title | Review_content | Traveller_type | Class |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Paige Boet | Jun-23 | New Orleans to London | 1.0 | Trip Verified | The airline lost my luggage | The airline lost my luggage and was absolutely | Solo Leisure | Economy Class |
| 1 | S Layne | Mar-23 | London to Amman | 1.0 | Trip Verified | fully refunded by our travel insurance | We booked on the BA website, round trip flight | Couple Leisure | Business Class |
| 2 | E Lanewoski | Heathrow to Bodrum | Business Class | 2.0 | Trip Verified | no boarding drinks provided | First time flying with BA business class, neve | A321 neo | Solo Leisure |
| 3 | Joel Burman | Jun-23 | Amman to London | 4.0 | Not Verified | WiFi didn't work | You can buy sandwiches and crisps but don't ex | Solo Leisure | Economy Class |
| 4 | R Vines | London City to Ibiza | Business Class | 7.0 | Trip Verified | stick with economy | This is a two-for-one review covering economy | Embraer 190 | Family Leisure |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3575 | W Benson | NaN | NaN | 4.0 | NaN | British Airways customer review | LHR-HKG on Boeing 747 - 23/08/12. Much has bee | Economy Class | no |
| 3576 | S Lugman | NaN | NaN | 4.0 | NaN | British Airways customer review | Just got back from Bridgetown Barbados flying | Economy Class | no |
| 3577 | D Smith | NaN | NaN | 4.0 | NaN | British Airways customer review | LHR-JFK-LAX-LHR. Check in was ok apart from be | Economy Class | no |
| 3578 | W Benson | NaN | NaN | 6.0 | NaN | British Airways customer review | HKG-LHR in New Club World on Boeing 777-300 | Business Class | yes |
| 3579 | Michael Dobbsen | NaN | NaN | 8.0 | NaN | British Airways customer review | YYZ to LHR - July 2012 - I flew overnight in p | Premium Economy | yes |

3580 rows × 9 columns

After this have printed the information about the dataset using the info() function.

```
# Printing information about the dataset
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3580 entries, 0 to 3579
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Passanger_Name   3580 non-null   object 
 1   Flying_month     2815 non-null   object 
 2   Route            2816 non-null   object 
 3   Rating           3575 non-null   float64
 4   Verified         1270 non-null   object 
 5   Review_title     3580 non-null   object 
 6   Review_content   3580 non-null   object 
 7   Traveller_type   3580 non-null   object 
 8   Class            3579 non-null   object 
dtypes: float64(1), object(8)
memory usage: 251.8+ KB
```

Printed the shape of the dataset which we got as 3580 rows and 9 columns.

```
df.shape
```

```
(3580, 9)
```

After this, we have the Statistical Summary of the Numerical column

```
# Statistical summary for numerical columns
numerical_summary = df.describe()
print("Statistical Summary for Numerical Columns:\n")
numerical_summary
```

```
Statistical Summary for Numerical Columns:
```

| | Rating |
|---|---|
| count | 3575.000000 |
| mean | 4.790490 |
| std | 3.170323 |
| min | 1.000000 |
| 25% | 2.000000 |
| 50% | 4.000000 |
| 75% | 8.000000 |
| max | 10.000000 |

And then a statistical summary for the categorical column.

```
# Statistical summary for categorical columns
categorical_summary = df.describe(include=['object'])
print("\nStatistical Summary for Categorical Columns:\n")
categorical_summary
```

```
Statistical Summary for Categorical Columns:
```

| | Passanger_Name | Flying_month | Route | Verified | Review_title | Review_content | Traveller_type | Class |
|---|---|---|---|---|---|---|---|---|
| count | 3580 | 2815 | 2816 | 1270 | 3580 | 3580 | 3580 | 3579 |
| unique | 2764 | 1174 | 744 | 2 | 2570 | 3508 | 203 | 13 |
| top | David Ellis | Aug-15 | Economy Class | Trip Verified | British Airways customer review | I really do not have the energy to write very ... | A320 | Economy Class |
| freq | 44 | 25 | 848 | 1067 | 952 | 2 | 342 | 676 |

Checking for null values in the dataset

```
# Checking if there are any null values in the dataset
df.isnull().sum()
```

```
Passanger_Name       0
Flying_month       765
Route              764
Rating               5
Verified          2310
Review_title         0
Review_content       0
Traveller_type       0
Class                1
dtype: int64
```

Handling null/missing values in the dataset

```
# Drop rows with missing values in 'Flying_month' and 'Route' columns
df.dropna(subset=['Flying_month', 'Route'], inplace=True)
```

```
# Impute missing values in 'Rating' column with median
median_rating = df['Rating'].median()
df['Rating'].fillna(median_rating, inplace=True)
```

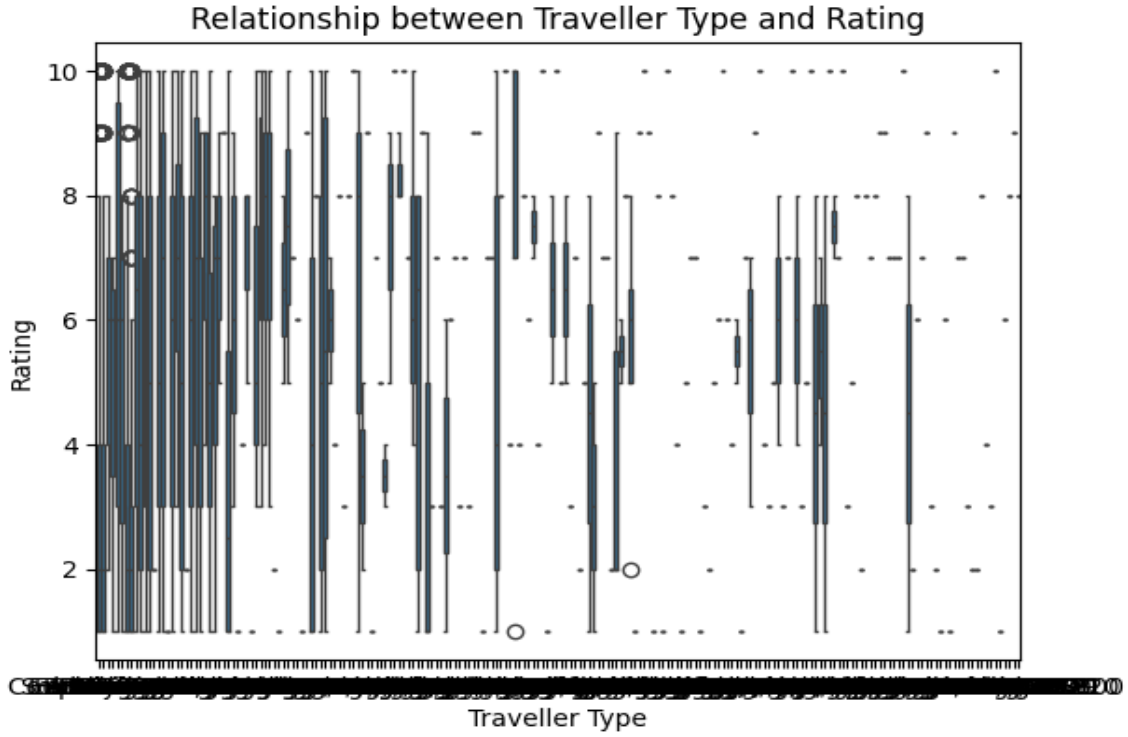
```
# Impute missing value in 'Class' column with most frequent class
most_frequent_class = df['Class'].mode()[0]
df['Class'].fillna(most_frequent_class, inplace=True)
```

```
# Replace missing values in 'Verified' column with 'Unknown'
df['Verified'].fillna('Unknown', inplace=True)
```

After handling the missing values/ null values in the dataset. Now doing Research questions analysis.

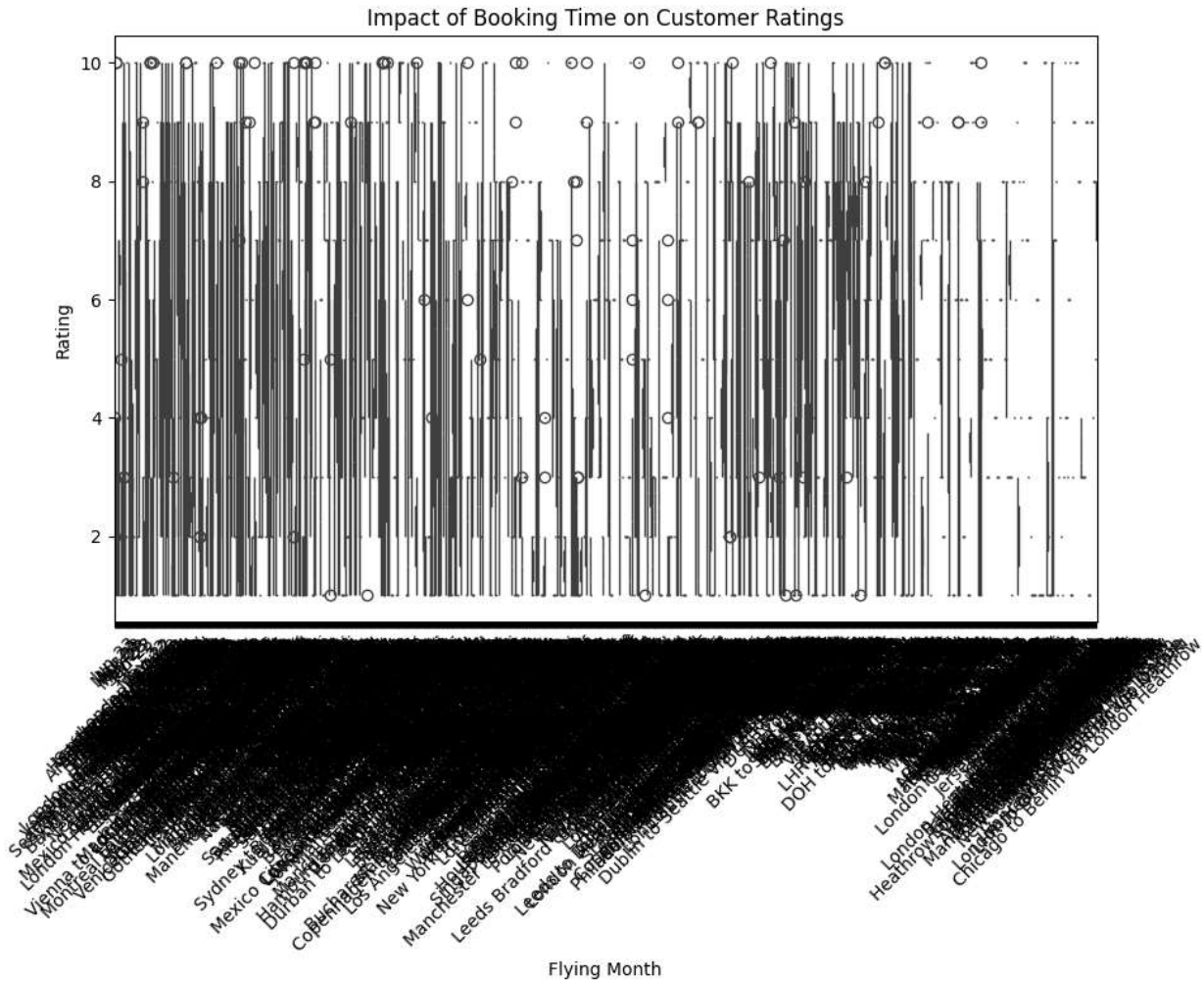
For the 1st research question i.e. "**What aspects of planning a vacation on an airline affect customers' decisions?"**

First have plotted the boxplot to examine the relationship between rating and traveller_type



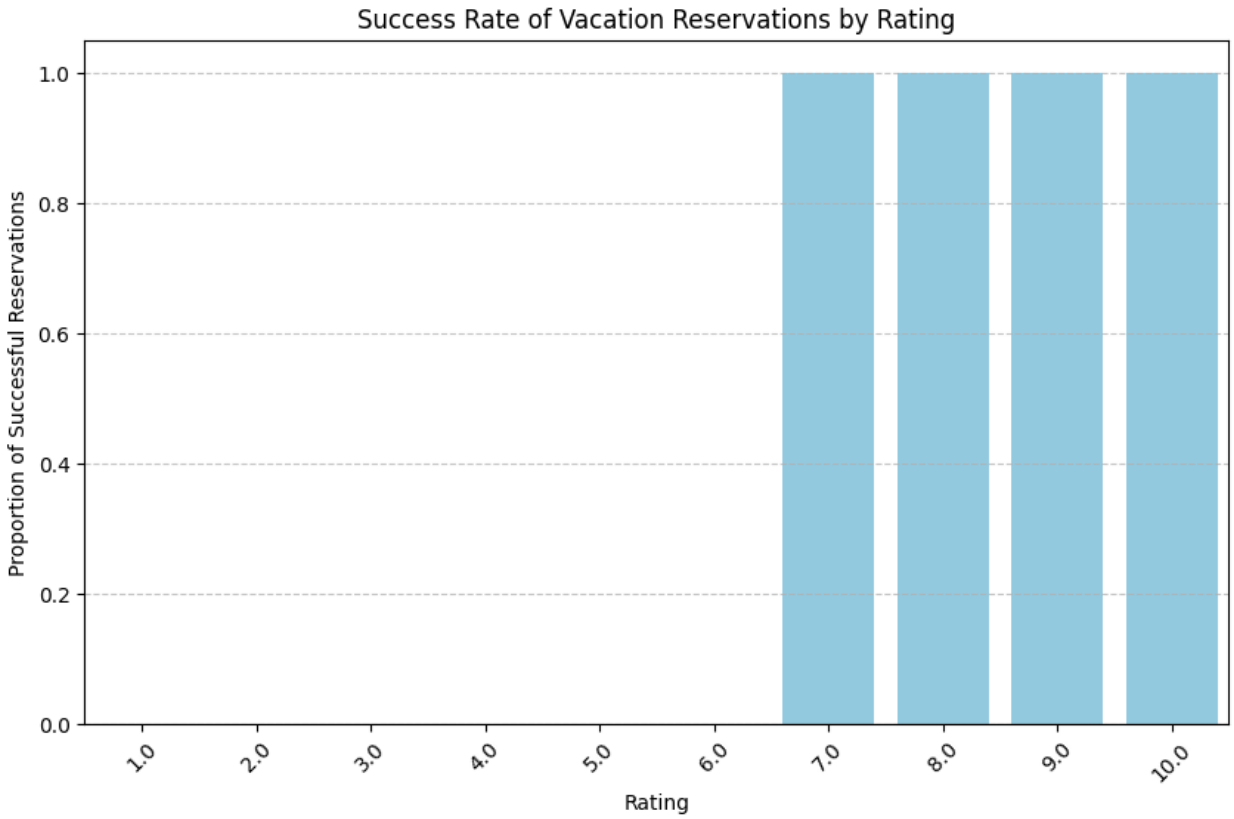

With your given boxplot, you are able to delineate how for a given there is a positive correlation between traveller type and rating. Such matter indicates that the people who put it as the main criterion ("The rating" or so) when making a choice of a carrier, also give more percentages for this rating.

For research question 2 i.e. " **What impact does the time of booking have on the preferences and behaviour of customers?",** have plotted the boxplot for comparing the distribution of ratings for different booking months.




In the graphic below is shown the impact of a time window for booking flights on the proper assessment of the customer. On the abscissa, we have the booking month plotted. On the ordinate - the customer rating. The line graph features alternating bars which signify the corresponding month where bookings were made. The middle 50% of ratings are marked with a central horizontal line which sits through the box; the whiskers, however, just refer to the high and low ratings which aren't any outliers.

For research question 3 i.e." **Based on the traits of the consumer, is it possible to forecast the chance of a successful vacation reservation?"** have plotted the bar plot for Success Rate of Vacation Reservation by Rating.

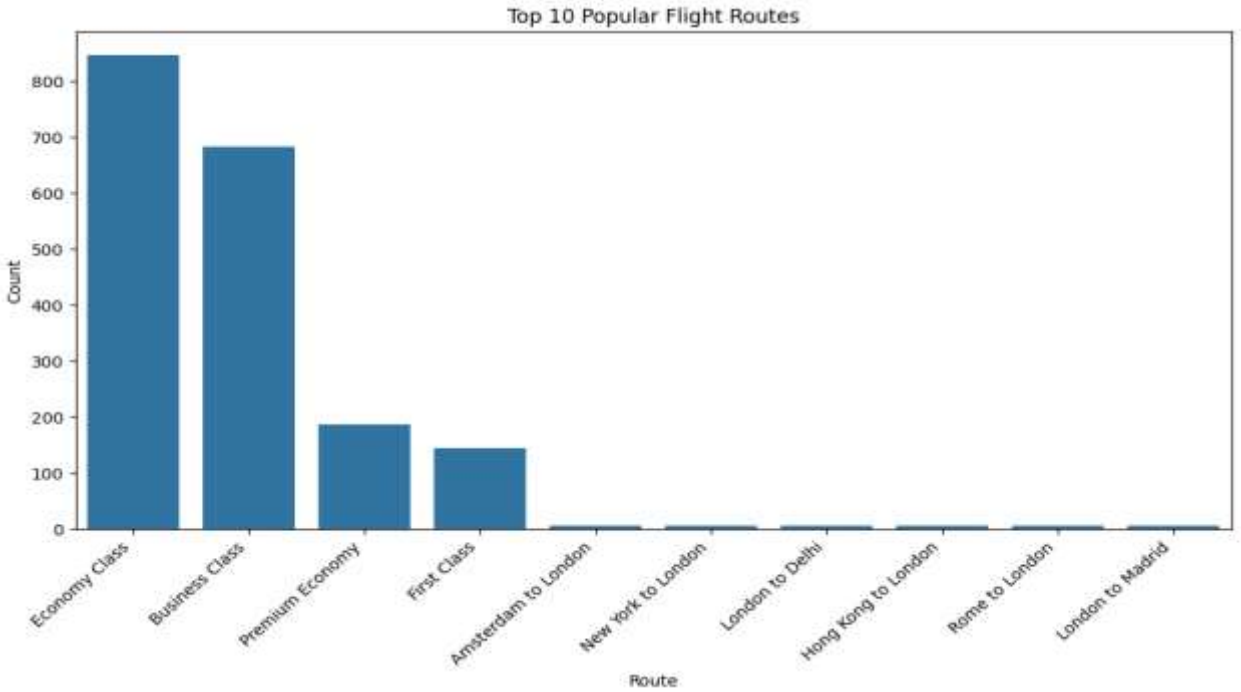



In the vacation reservation of rate the by bar of plot it mean that, it represents that resort can't tell the chance of reservation success due to case characteristics. Booking time, location, passanger type and betgem might be just a few factors that influence a reservation to be booked or cancelled. And again, the chances of reservation success are erratic, because many people rate either their total experience or the qualities of the place, on a different scale, such as the facility, entertainment, affordability, and time deadline. Thereby, as in the case of the map, the forecast will not understand all the details.

For research question 4 i.e." **Which flight schedules and itineraries are the most popular among passengers?".** Here we have 10 popular Flight Routes:

```
Popularity of Routes:
 Route
Economy Class            846
Business Class           683
Premium Economy          186
First Class              143
Amsterdam to London        6
New York to London         6
London to Delhi            5
Hong Kong to London        5
Rome to London             5
London to Madrid           5
Name: count, dtype: int64
```

And have also plotted the countplot Top 10 popular Flight Routes




Looking into its own weighted customer base, airlines may find gap in markets as they have the capability to get the highest in-demand and under-reserved travel routes with the help of the count plot, which highlights the top 10 most chosen flights based on attendance numbers. The most prevalent routes are between Amsterdam - London, New York - London, and Delhi - London, showcasing a considerable need of flying between well-known business centres. The popular cities with the largest number of leisure travelers' demand are Madrid and Rome in Europe.

Performing the Chi-square test of independence to test the hypothesis that customer attributes influence the chance of a successful vacation reservation.

```
from scipy.stats import chi2_contingency

# Create a contingency table for traveller type and reservation status
contingency_table = pd.crosstab(df['Traveller_type'], df['Reservation_Status'])

# Print the contingency table
print("Contingency Table:")
contingency_table
```

Contingency Table:

| Traveller_type | Reservation_Status Successful | Unsuccessful |
|---|---|---|
| 767-300 | 0 | 1 |
| 777 | 1 | 2 |
| 777-200 | 0 | 1 |
| 777-300 | 1 | 0 |
| 787 | 0 | 1 |
| ... | ... | ... |
| SAAB 2000 | 1 | 0 |
| Saab 2000 | 1 | 0 |
| Solo Leisure | 55 | 252 |
| Various | 0 | 1 |
| boeing 787 | 0 | 1 |

```
# Perform the chi-square test of independence
chi2, p_value, dof, expected = chi2_contingency(contingency_table)

# Print the test results
print("\nChi-square Test Results:")
print(f"Chi-square statistic: {chi2}")
print(f"P-value: {p_value}")
print(f"Degrees of Freedom: {dof}")
print("Expected frequencies:")
print(expected)
```

```
Chi-square Test Results:
Chi-square statistic: 458.4828700307012
P-value: 1.5863250173013544e-22
Degrees of Freedom: 199
Expected frequencies:
[[  0.32326821   0.67673179]
 [  0.96980462   2.03019538]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  1.61634103   3.38365897]
 [ 35.23623446  73.76376554]
 [  0.64653641   1.35346359]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.96980462   2.03019538]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]
 [  0.32326821   0.67673179]]
```

```
# Interpret the test results
alpha = 0.05
print("\nHypothesis Testing:")
if p_value < alpha:
    print("Reject the null hypothesis (H0). There is a significant association between customer attributes and reservation status.")
else:
    print("Fail to reject the null hypothesis (H0). There is no significant association between customer attributes and reservation status.")
```

```
Hypothesis Testing:
Reject the null hypothesis (H0). There is a significant association between customer attributes and reservation status.
```

**Interpretation of the results:**

**Contingency Table:** One of the major functions of a contingency table is to allow us to determine the correlation factor between the two categorical variables, i.e. reservation status and traveller type. Every combination of traveller is present resulting in the success and failure of the reservation details.

**Chi-Square Test Results:**

**Chi-Square Statistic:** The number of bars is determined by converting the calculated value of the chi-square statistic into corresponding bars.

**P-value:** This value is equal to 1.58e-22 and corresponds to the p-value of the chi-square statistic.

**Degrees of Freedom:** The degree of freedom for the test are 199.

**Expected Frequencies:** Those frequencies are estimated onwards when the data is analyzed using the variable settings of their independence which are referred to as the expexted frequencies.

**Interpretation:**

**Null Hypothesis(H0):** In line with the innocent hypothesis, the independent variable and the customer variables such as the kind of the traveller do not appear to be significantly correlated.

**Alternate Hypothesis(H1):** Alternatively, the research hypothesis suggests that the association between the consumer variables and the non-white respondents level is statically significant.

**Conclusion:**

- We gladly accept the alternative hypothesis since the observed p-value (p = 1.59e-22) is less than the set significance level (α = 0.05).
- Therefore, we draw the conclusion that reservation status and the identity of travelers (travelers' type) are significantly dependent on each other.

**Task 4: Result Evaluation and Future Development**

Data analysis of the airline review data set reveals a variety of passenger choices and experiences; this disclosure of facts is beneficial to the non-users o f air castle. These revelations may have the following possible effects on the aviation industry's application domain:

1. **Improving Customer Satisfaction:** Airlines can became more consumer oriented by developing such marketing strategies as personalized service that uses the knowledge of predictive factors that influence customers decisions and actions. Airlines can do this by reviewing customer preferences and demands

and determining where the bulk of the passengers are along with the mostly preferred flight schedules and itineraries.

2. **Enhancing Service Offerings:** Aviation lines may be able to see areas where they can improve or areas that are strengths based on guest comments and ratings. With the aid of this data customers can be provided with preference of cabins containing comfortable seats, more efficient catering and significant improvement of flight attendants attitude.
3. **Marketing and Customer Targeting:** Customer segmentation to be enhanced through using the gathered data like client demographics and booking patterns for better-targeted marketing strategies. Airlines could shoulder some efforts in raising consumer engagement and retention by analyzing these data for marketing purposes and personalizing it into special deals and promotions that suit different customer segments.
4. **Operational Efficiency:** While airline companies may improve resource allocation and speed up the booking processes by knowing what types of travellers and booking time can influence interest to book, it is important to understand that these predictions can rarely provide direct insights into the variation in demand across airlines. Such an outcome can help to receive cost and enhance the efficiency of the business operations.

**Limitations of the Analysis:**

1. **Data Quality and Completeness:** The authenticity and totality of the data set become urgently important prerequisites for the study. Bad or deficient data that we use in the analysis might endanger the results' believed accuracy.
2. **Sample Bias:** The fact that the dataset may not accurately capture everybody who traveled by airline plane personally is also likely. Concerns on the biasness of reviews' selection or reporting and this might skew the results and thus make the findings quite narrow.
3. **Causality vs. Correlation:** While following can lead researchers to correlations among various features, proving cause or effect cannot be done by the use of research. The consumer's preferences and behaviours are further affected by variables that are beyond the dataset like social media influences or personal experiences.

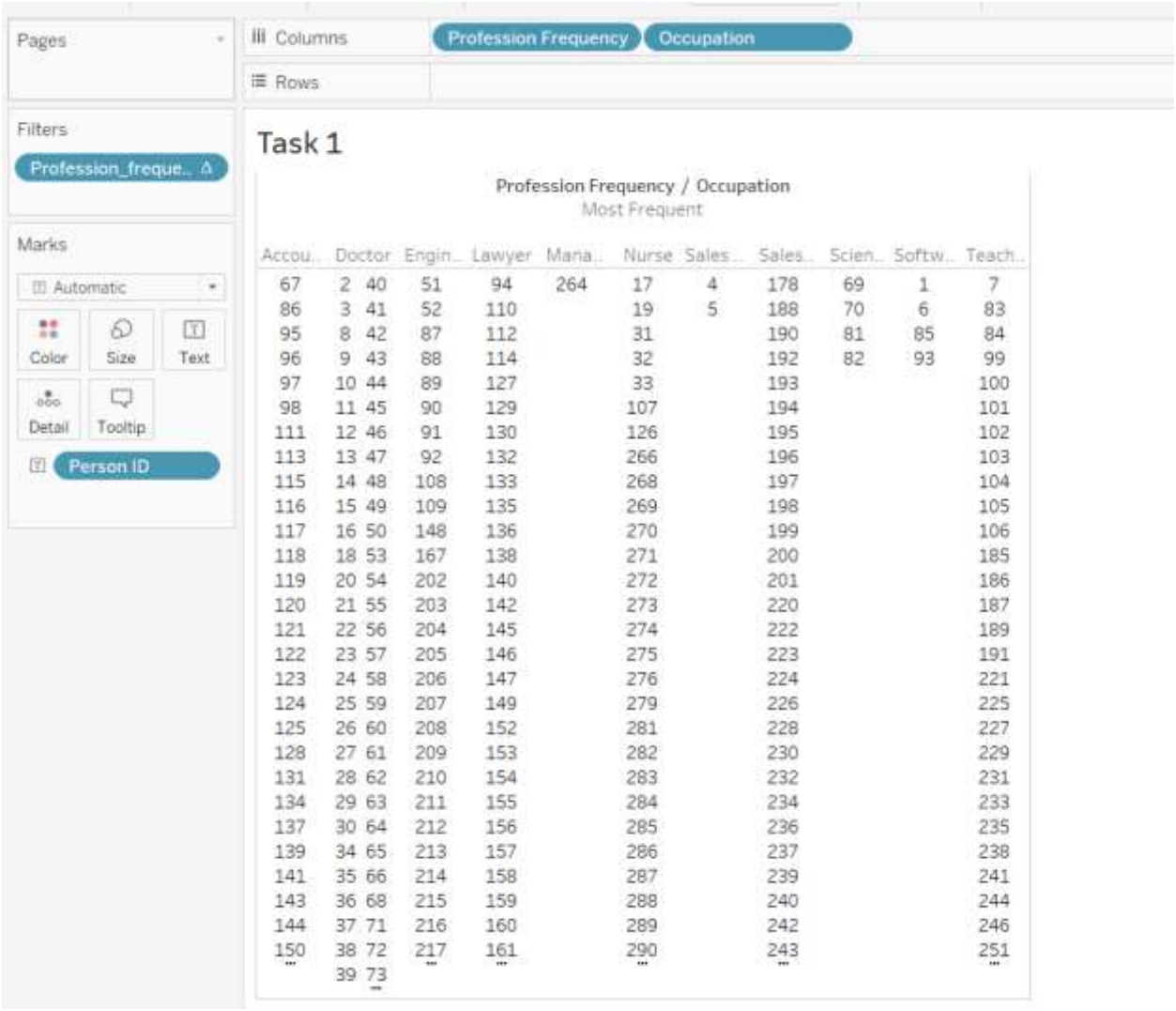
**Future Work:**

1. **Sentiment Analysis:** Emphasis on knowing the attitude and feeling of the consumers as they try airline services can be achieved through implementational of natural language processing and sentiment analysis on their actual descriptions of the services.
2. **Predictive Modeling:** Eventually, the operations of the low-cost airlines can be based on the predictive models that anticipate the actions of the customers, i.e. changes their travel selections or cancel their reservations.
3. **Real-Time Data Analytics:** Airlines possibly will catch up and learn how to analyze and interpret information almost instantaneously by employing machine learning and AI features. It would permit to make up the for any gaps and adjust the services, when and how needed.
4. **Integration with Customer Relationship Management (CRM) Systems:** The profiling results, if combined in a way with CRM systems could help in tailor-making marketing campaigns that are based on individual preferences and behaviour and also create engagement with customers in a personalized way.

Eventually, the solution will hopefully be able to add extra profitability to the aviation industry by being more successful in terms of customer satisfaction, operational efficiency and general business success, taking all these factors into account as well as the introduction of innovations in technology and analytics.

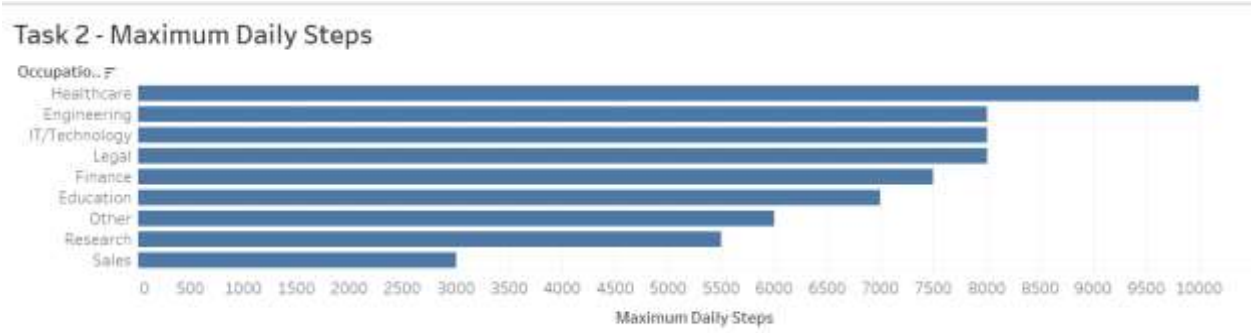
**Section 2 - Business Intelligence ( Tableau )**

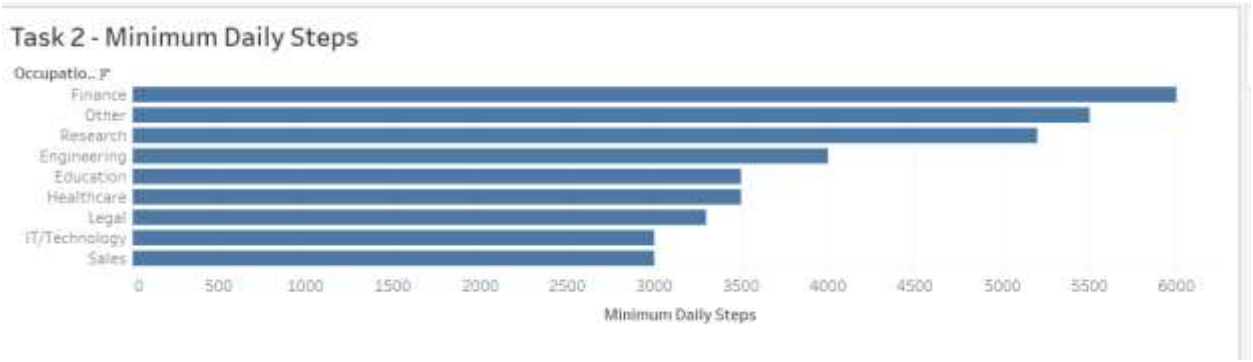
**Task 1**




The data presented in the blood pressure visualization filters data and spotlights atypical blood pressure in occupations using the 'Profession Frequency' parameter. The "Profession Frequency" field limits the records that are abnormally high in blood pressure, while the "Person ID" field merely shows the number of entries that are available for the group of occupations. The users are allowed to access degree of abnormality of the different professions with hypertension issue either the most severely afflicted or the least through the dynamic displaying of the occupations using Tableau. The understandable interface, sure enough, covers the detailed representation of work-related regular blood pressure variations among the listed positions. The parameter capability of the Tableau can be made more flexible. This can so happen so that the people who are stakeholder may know how the job and the blood pressure are related.

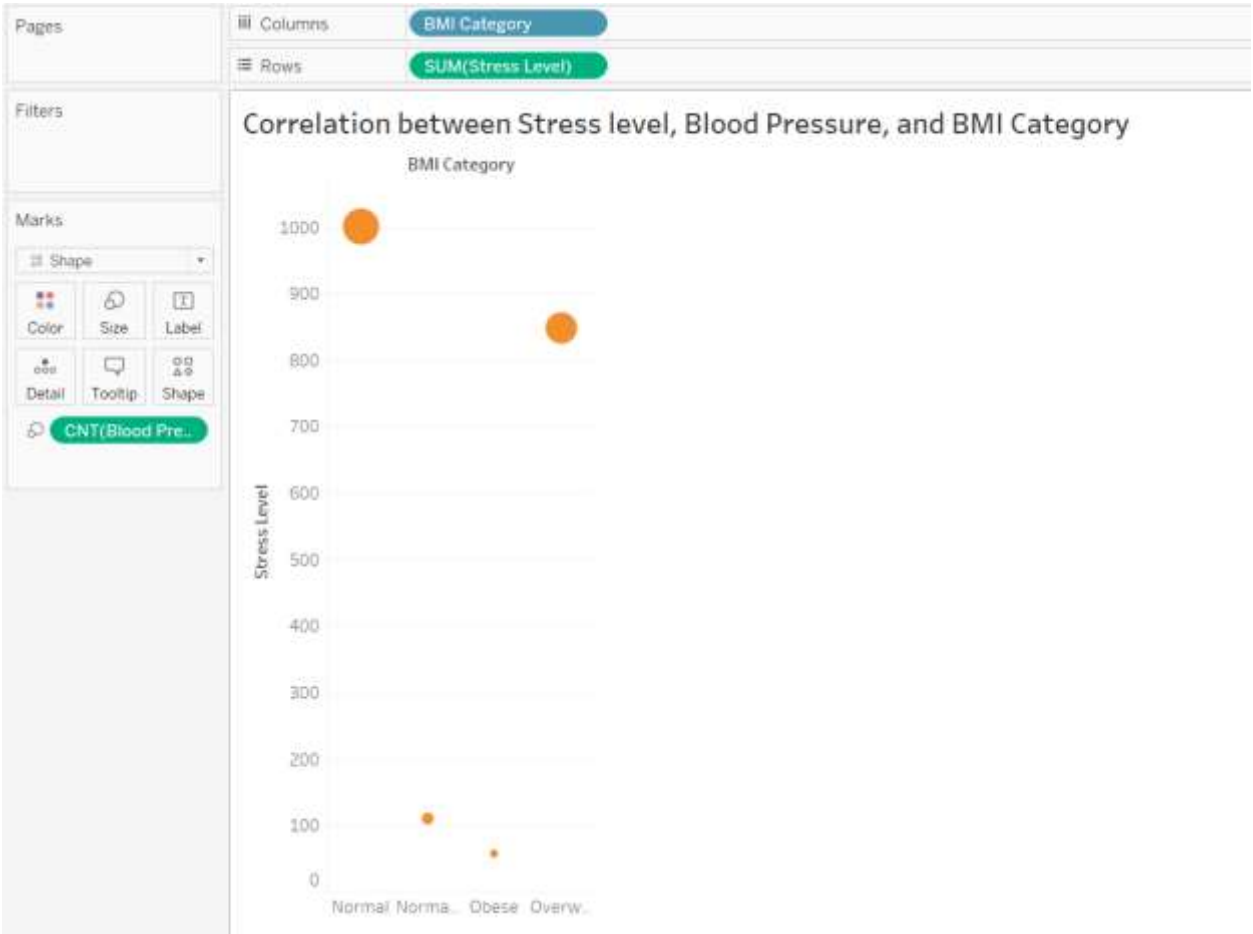
**Task 2:**






With a combination of a morning rush, a midday lull, and an evening bump, some workers appeared to take more steps during certain hours while most of them had their own routine. The bar chart indicating consecutive steps per job type with the biggest and minimal steps - in descending order, was the visualization shown. To create a more general grouping and for the analysis of each category, the "Occupation Category" field was created for roles such as IT/Technology, Healthcare, Sales, and Education. To demonstrate the range between the highest and the lowest step counts the originally computed fields such as “MADAYSTEP”and “MINDAYSTEP” were listed in columns and were recollocated in descending order. The ensuing bar chart, where bars stand for the types of work and their widths represent the variation in the number of steps taken per day, is a comparatively simple graphic to understand. The pictorial representation of the ranges daily steps by days, highlights the readability of the visualization and gives a quick grasp opportunity of the data by the public.

**Task 3:**



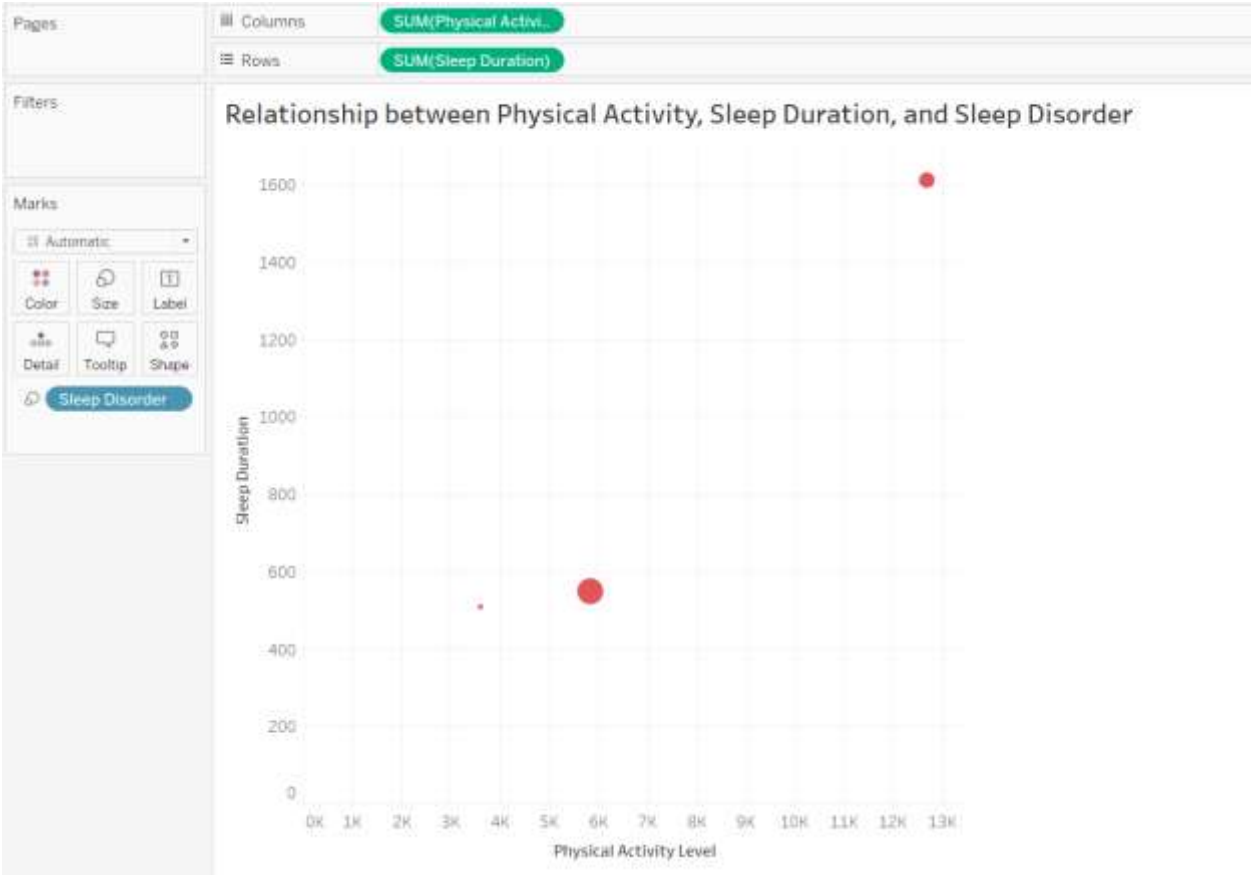

Task3 generated graph representing BMI and stress levels versus blood pressure was done usingTableau. To assess any relation between these influential variables as for health conditions a scatter plot was determined. Data such as BMI categories, stress levels and blood pressure came out being organised in columns, rows and sizes. A data point represented a subject by stress level, BMI and blood pressure on the given figure. The existence of these relations was known to follow through from observing the location of the data points on each axis.

Stakeholder were able to spotline the trends or patterns between blood pressure, the level of stress, and the BMI through the scatter plot analysis. On the other hand, the classifications that emerged in the form of pooling of points and trends cutting across at various categories and levels of stress suggested certain links, relationships, or patterns, indicating possibilities.

Learning about the data points on a scatter plot may bring an assurance that these things may not only be independent from each other but as well interact with one another.

In short, the scatter plot animation provides the detailed outlook for a possible connection between blood pressure, tension degree, and BMI. This knowledge will be able to be something that improves future study. This will also have an influence on the increased health and wellbeing.

**Task 4**

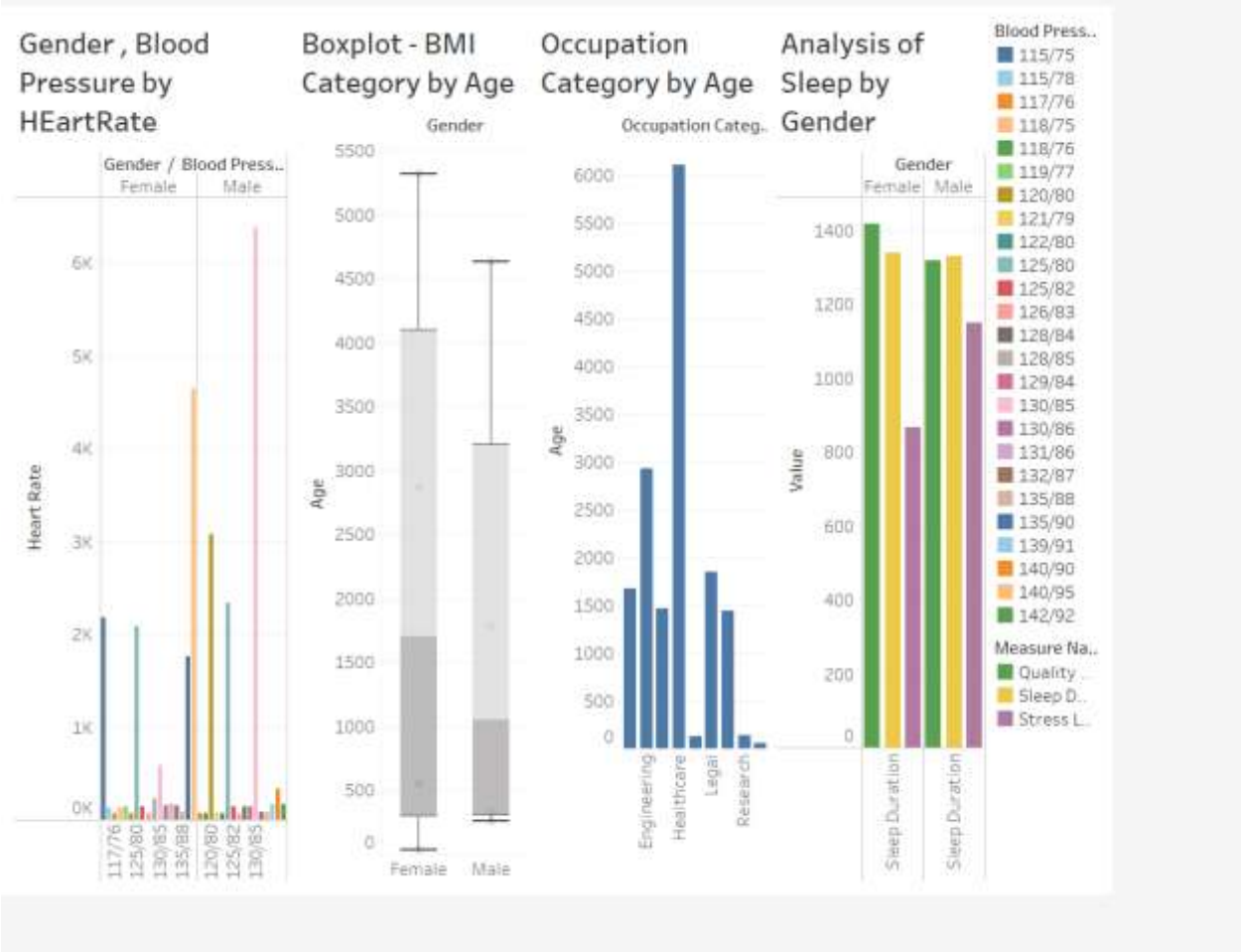



Power of Task 4, conducted in the frame of Tableau software, was used to explore the relation between expenditure of energy, number of sleeping hours and sleep problems. It was depicted on a scatter plot to show the relationship between the physical activity degree and the length of sleep, as well as the presence or absence of a sleep disorder. Each dot in the figure was assigned for a person. Further investigations might be made from such trends that can be inferred from the relative placement of data points at each axis.

By using the scatter plot, the stakeholders will easily spot patterns and correlations with hours of exercise and amount of sleep with insomnia. Broad relationships or connections between the development of sleep disorders may be displayed by geometric figures such as clustering, trends and continuous lines spanning from physical activity to sleeping duration.
The study of the spread line data can help one to understand in depth the relationships and coexistence among physical activity, sleep length, sleeping disorders.

The scatter plot visualization done acts not only to educate stakeholders on how sleep health may be affected by the lack of physical activity, length of sleep, and different sleep disorders. It also aids in the development of intervention systems that could be used to enhance the level of quality of sleep, the duration of sleep, and overall well being.

**Task 5**




The objective of Task 5 was to build four interactive visualisations which would help to perform the data investigation in different directions. By presenting related data on various pages and featuring dynamic data interaction, the dashboard takes a step further towards allowing even better navigation and data control to be performed by the users. The diagram represents age as a basic flow, the pressure of blood rate by heart rate, and the sleep of women and men, as well as also plots the card of BMI by age.

All the graphs show the parameter differences of sleep between genders such as sleep time, sleep efficiency, total movement, and wake after sleep. While clicking on any chosen data points or categories enclosed within a single visualization, users could additionally interact with the dashboard components, evaluating covariations between variables and looking for connections, patterns or tendencies in the process.

Organised data-information exchange and decision-making process becomes possible with the responsible approach to the contemporary TV-viewing. First of all, after processing a dataset the dashboard presents a detailed of the information and lets user directly drill into the similarity, trends, and distribution within the data.

The figure is called “Labor Category by Age” and it as it looks as a bar chart. On the chart, the horizontal axis shows different age groups while the vertical axis displays the numbers of people in every labor category.

The age distribution charts by occupational category are good resources that one uses to know the make-up of the workforce. They can show things like:They can show things like:

Should we target young age groups or a general population group for specific occupations?
If it applies to a group of older people or younger people, this will be the type of employment or job.Whether this re-arrangement of workers among different fields will take place and how the future distribution varies from the current one.

The tabulated data in the picture present the blood pressure values for different genders expressed at different heart rates. On the negative side, excessive human measurements result in missing pulses, lack of blood pressure measurement (diastolic or systolic), and increased age that significantly influences atrial depolarization. Men substantially have a high blood pressure than women, in particular aging men, but women are on are little more rapid heartbeat. The link between heart disease and risks like smoking, obesity, cardio inactivity, and family history have been established. Only when all relevant parameters are captured and investigated in a thorough way can valid findings be reached going beyond the available evidence. Beyond this shortcoming, this picture leads to an

understanding of the inequalities of both genders in the traits, specifically, heart rate, blood pressure, and pulse rate.

The "Analysis of Sleep by Gender" chart includes men's and women's average length, quality and stress level within sleep. Women's sleep is longer than that of men's, but women are more likely have a poor-quality sleep due to excessive stress. The table exhibits that the slumber routine amongst adolescents differs from either grownups or young children, with the latter ones going to bed later and waking up late than their age-mates.

**Conclusion:**
A detailed understanding of airline customers' satisfaction and healthcare aspects could be derived applying both Tableau data visualization and Python statistics analysis. Airlines use Python to obtain such data as guest evaluations in order to come up with a new path or strategy this is according to the passenger wants and behaviours. Airlines can create a positive experience among customers and increase their loyalty by identifying feedback patterns obtained from users and through big data analytics such as sentiment analysis and classification algorithms.

On the other hand, Tableau visualisations permit to study the relations between fitness indicators like arterial pressure, body mass, and stress level, as well as physical activity, the sleep time, and sleeping disorders. Through the implementation of those information, the chance for airlines to make targeted measures on what will improve passengers' experience and overall well-being is increased. Health issues and customer satisfaction are gaining significance in tough aviation sector; airlines can look into these concerns and can improve the feel of long-term loyalty and rewards.