ANTISOCIAL BEHAVIOUR ANALYSIS

Table of Contents

Section 1: Big Data Analytics (Python) – Analyzing antisocial behavior in London using the dataset *MPS_Antisocial_Behaviour.csv*.

Introduction

Antisocial behaviour is a serious issue that affects urban areas, leading to the increase in safety concerns and a decline in the quality of life among residents. The Metropolitan Police classify ASB incidents into three categories: Personal, Nuisance, and Environmental. The dataset, MPS_Antisocial_Behaviour.csv, records ASB incidents from August 2023 to August 2024, providing details on type, location, and time. Proper analysis of the data is important in understanding the patterns, identifying high-risk areas and implementing better policing strategies that will help in dealing with the problem in the best way possible.

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

State-of-the-Art Literature in Antisocial Behaviour (ASB) Analysis

The study of antisocial behaviour (ASB) has been instrumental in criminology, data science and urban planning. Recent advancements are taking advantage of Big Data, Machine Learning and Geographic information system (GIS) to analyse ASB trends and predict crime hotspots. Viding et al., (2024) reveal that the origins of antisocial behaviour are complex and vary between individuals. The authors appreciate the fact that ASB have makes individuals and society to incur significant costs. According to Viding et al., (2024) ASB includes rule breaking and aggression that originates from childhood and may be associated with several mental health problems, special education needs, legal system involvement, physical health problems as well as failure to complete education. In addition to the financial implications, ASB also has several social and emotional costs that can not be measured in financial terms. It is therefore important to find lasting solutions to the challenge as a way of improving the quality of life.

People should be taught about the implications of antisocial behaviours as a way of improving the quality of life in society. Esposito et al., (2021) argue that adolescent thinking about the wrongness of ASB is important in making in preventing them from taking part in such behaviours. However, the measures that can be used in assessing adolescent antisocial behaviours are not common in literature despite the fact that the field is important. Esposito et al., (2021) therefore developed and validated a new dimensional scale of antisocial behaviour evaluation. Such evaluation will help in understanding how adolescents perceive antisocial behaviours, which will make it possible to come up with the best ways to deal with ASB. Adolescents can also use the tool for self evaluation, which will make them know the challenges associated with ASB hence help in fighting it.

Application Domain of the Dataset

The MPS_Antisocial_Behaviour.csv dataset belongs to the domain of urban crime analysis, law enforcement, and public safety management. The dataset is primarily important in understanding and mitigating antisocial behaviours (ASB) in metropolitan areas. The dataset has several key areas of application that include law enforcement and crime prevention, urban planning and public policy, big data analytics and AI modelling, as well sociological and behavioural studies.

Description of the dataset

The dataset is made of 255,432 records with 27 columns, capturing antisocial behaviour (ASB) incidents reported in London. The key features include date and hour, incident type, location data (Opening_Type_1, Opening_Type_2, Opening_Type_3), Location Data (Ward, Borough), response time, ASB count as well as safer neighbourhood team details. Date & hour includes the time stamp of the reported incident, incident type (Opening_Type_1, Opening_Type_2, Opening_Type_3) includes the classification of ASB such as rowdy behaviour, drug related and noise.

```
=== DATASET OVERVIEW ===

Shape of dataset: (255432, 27)

First 5 rows:
```

	Date	Hour	OP01	Opening_Type_1	OP02	Opening_Type_2	OP03	Opening_Type_3	CL01	Close_Type_1	CL02	Close_Type_2	CL03	Close_Type_3	Resolution_Type
0	2024-04-07	19:00	215	ASB Nuisance	202	Rowdy Or Inconsiderate Behaviour		NaN	215	ASB Nuisance		NaN		NaN	Police Attendan Not Requir
1	2024-04-01	20:00	215	ASB Nuisance	601	Drug Related		NaN	215	ASB Nuisance		NaN		NaN	Inform / Inform
2	2024-04-14	16:00	215	ASB Nuisance	202	Rowdy Or Inconsiderate Behaviour	601	Drug Related	215	ASB Nuisance		NaN		NaN	Link
3	2024-04-04	18:00	215	ASB Nuisance	202	Rowdy Or Inconsiderate Behaviour		NaN	215	ASB Nuisance		NaN		NaN	Inform / Inform
4	2024-04-02	18:00	215	ASB Nuisance	204	Rowdy / Nuisance Neighbours	211	Noise	215	ASB Nuisance		NaN		NaN	Inform / Inform

```
Column names and data types:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 255432 entries, 0 to 255431
Columns: 27 entries, Date to Safer_Neighborhood_Team_Borough_Code
dtypes: float64(1), int64(4), object(22)
memory usage: 52.6+ MB
None
```

Figure 1: The first five rows of the dataset and the datatypes contained

Location data (ward, Borough) is the geographic area where the ASB was reported. Response time is the time taken to respond to the incident. ASB count represents the number of incidents in a given report. Safer Neighbourhood team details represent the team handling ASB in specific boroughs. The dataset will help in identifying high-risk areas for ASB, understanding time-based trends of ASB occurrences, and assisting police departments in strategic resource allocation.

Summary statistics:

	Date	Hour	OP01	Opening_Type_1	OP02	Opening_Type_2	OP03	Opening_Type_3	CL01	Close_Type_1	CL02	Close_Type_2	CL03	Close_Ty
count	255432	255432	255432.00	255432	255016	251538	243850	112675	252785	252366	235010	1982	234876	
unique	366	24	NaN	4	116	114	136	133	59	58	46	45	18	
top	2023-10-31	19:00	NaN	ASB Nuisance	202	Rowdy Or Inconsiderate Behaviour		Drug Related	215	ASB Nuisance		ASB Nuisance		Unli C
freq	1406	16973	NaN	219497	145571	145571	131174	35693	170914	170914	233028	450	234849	
mean	NaN	NaN	214.92	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	2.17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	11.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	215.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	215.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	215.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	216.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Figure 2: Summary statistics of the data

Research questions

1. What are the most frequent types of antisocial behaviour (ASB) in London, and how do they vary across different boroughs?
2. What time of day and which days of the week experience the highest ASB incidents?
3. Does ASB incidents correlate with socioeconomic factors such as population density or income levels in different boroughs?
4. How do seasonal variations (summer vs. winter) impact the frequency and nature of ASB incidents?

Hypothesis formulation

Null Hypothesis (H_0): The distribution of ASB incidents is uniform across all boroughs in London.

Alternative Hypothesis (H_1): Certain boroughs have significantly higher ASB incidents compared to others.

Statistical Test: Chi-Square Test for independence to examine borough-wise ASB distribution.

Task 2: Evaluation of Approaches and Technologies for Big Data Analysis of Antisocial Behaviour (ASB)

Evaluating various approaches and technologies for performing the data evaluation tasks and developing Big Data applications.

Robust Big Data techniques are required to handle large datasets in the best way possible when analysing antisocial behaviour (ASB). A mixture of traditional statistics methods was used for general understanding of the data. Machine learning was used for predictive analytics. The traditional statistics methods were used to provide descriptive statistics correlation analysis and hypothesis testing. The analysis was important in summarizing ASB trends. The technologies used in traditional statistics methods include Pandas, NumPy, SciPy libraries. Machine learning was used to predict ASB hotspots using classification and regression models. They were also used to identify patterns in ASB based on time, location and socio-economic data. The machine learning models that were used in this case include Decision Trees and Random Forests for structured data like ASB reports and K-Means Clustering for grouping the ASB prone areas. The technologies used include Scikit-learn, TensorFlow, XGBoost, which are important machine learning models.

Solutions and Techniques Applied to Similar Problems

Machine learning is important in the identification of antisocial behaviour in the social media networks. Singh et al., (2023) researched on how machine learning models can be used in identification of antisocial behavior on Twitter hence making the platform safe for the users. According to the authors, online antisocial behavior is a social problem that can negatively affect public health hence it should be dealt with in the best way possible.

Law enforcement agents need to know the ASB hotspots so that they can deal with the problem appropriately. Murray (2022) reveals that police can easily fight crime if they are able to accurately predict where and when a future crime is likely to happen. The author suggests near repeat analysis, and hotspot mapping some important techniques that can be used to identify and fight crimes. Machine learning technologies can be used to fight crimes such as delinquency and criminal behavior according to Schroeders et al., (2023). The information on how machine learning has been used to fight crime in the past is important because it shows how the same technologies can be used to fight numerous criminal behaviours such as ASB.

Chosen Methodological Approach and Justification

Python libraries such as Pandas, NumPy, and SciPy were used for data analysis. Moreover, Scikit-learn was used for machine learning. Mean imputation and filling the missing values with N/A was important in dealing with the missing values during the data cleaning process. The timestamps were converted into meaningful time-based features (Schroeders et al., 2023). The justification for the data cleaning process is that it ensured clean and structured data before analysis.

Exploratory data analysis was also used to ensure easier understanding of the dataset and make the necessary conclusions from it. Heatmaps and bar charts were used to visualize ASB trends. Chi-Square tests were used to assess ASB distribution by borough. The importance of exploratory data analysis is that it helped in the identification of key ASB patterns before predictive modelling.

Machine learning approaches were important for predicting crime trends to make it easier for the law enforcement agencies to fight crime in the best way possible. Random forest classification was used to predict boroughs with high ASB risk. Time-Series Forecasting (ARIMA) was also applied to predict future ASB incidents (Schroeders et al., 2023). The justification for the choices is the fact that Random Forest handles categorical ASB data well and the time series models make it possible to properly forecast ASB trends.

Approaches and Technologies for Data Evaluation & Big Data Applications

Data Preprocessing in Python

The main objective of the data preprocessing in Python is to clean, structure and prepare data for analysis. The first step was to handle the missing values and duplicates. The missing values were identified then then mean imputation was used to fill the missing space to ensure that the dataset is as consistent as possible. The Python code snippet was used to count the missing values per column.

```
import pandas as pd

# Load dataset
df = pd.read_csv("MPS_Antisocial_Behaviour.csv")

# Check for missing values
print(df.isnull().sum())  # Count missing values per column
```

Figure 3: Code to identify missing values

Running the code brought up the output below:

```
MAHIGI@DESKTOP-MMGJTJ1 MINGW64 ~/Documents/projects/data_analysis
$ ./asb_data_analysis.py
C:\Users\MAHIGI\Documents\projects\data_analysis\asb_data_analysis.py:6: DtypeWarning: Columns (13) have mixed types. Specify dtype option on i
mport or set low_memory=False.
  df = pd.read_csv("MPS_Antisocial_Behaviour.csv")
Date                                      0
Hour                                      0
OP01                                      0
Opening_Type_1                            0
OP02                                    416
Opening_Type_2                         3894
OP03                                  11582
Opening_Type_3                       142757
CL01                                   2647
Close_Type_1                           3066
CL02                                  20422
Close_Type_2                         253450
CL03                                  20556
Close_Type_3                         255405
Resolution_Type_1                      2715
Resolution_Type_2                    235160
Ward                                  20350
Ward_Code                            255432
Response_Time                             0
Duplicate                                 0
ASBCount                                  0
Datetime                                  0
Ward_WardCode                          2472
Safer_Neighborhood_Team_Name           2472
Safer_Neighborhood_Team_Code           2472
Safer_Neighborhood_Team_Borough_Name   2472
Safer_Neighborhood_Team_Borough_Code   2472
dtype: int64
```

Figure 4: The output after running the code above

The output showed that there were missing values in multiple columns.

The next step was to deal with the data type warnings. There was a datatype warning on column 13 (*Close_Type_3*), which had mixed types (numbers and text). It was important to convert it properly before dropping or filling the missing values using the code snippet below:

```
# Convert mixed-type columns to string
df['Close_Type_3'] = df['Close_Type_3'].astype(str)
```

Figure 5 Dealing with the mixed data types

Some of the columns were completely empty e.g. *Ward_Code*, *Close_Type_3.* Columns like Ward_Code (100% missing) and Close_Type_3 (99.99% missing) were dropped because they provided little or no useful information that would be helpful in the current research. The code below was used in dropping the less important columns.

```
# Drop columns with excessive missing data
df.drop(columns=['Ward_Code', 'Close_Type_3', 'Close_Type_2'], inplace=True)
```

Figure 6: Code to drop excessive missing data

I filled the missing categorical values such as Ward, Opening_Type_2, and Safer_Neighborhood_Team_Borough_Name with N/A, meaning that they were not provided using the code snippet below:

```
# Fill missing categorical values with "N/A"
categorical_cols = ['Ward', 'Opening_Type_2', 'Safer_Neighborhood_Team_Borough_Name']
df[categorical_cols] = df[categorical_cols].fillna("N/A")
```

Figure 7: Filling the missing values with N/A (not available)

The numerical columns such as Response_Time I used the mean because response times generally follow a normal distribution.

```
# Fill missing numerical values with the mean
df['Response_Time'].fillna(df['Response_Time'].mean(), inplace=True)
```

Figure 8: Filling the missing numerical values with the mean

The dataset did not have duplicate values but the following lines of code were added so that they can be used in the future datasets that could have duplicate entries.

```
# Remove duplicate records
df.drop_duplicates(inplace=True)
```

Figure 9: Code snippet to deal with duplicate entries

The final step in the data cleaning process was to conduct the final check after cleaning using the code below:

```
# Verify missing values are handled
print(df.isnull().sum())

# Confirm no duplicate records remain
print(f"Remaining duplicate rows: {df.duplicated().sum()}")
```

Figure 10: Conducting the final check after completing the data cleaning process

The next step in the data cleaning process was to convert the date and time to datetime format using the code below:

```
# Convert date and time to datetime format
df['Date'] = pd.to_datetime(df['Date'])
df['Hour'] = pd.to_numeric(df['Hour'], errors='coerce')
```

Figure 11: Convert date and time to datetime format

The data was not clean for analysis.

Descriptive Statistics & General Analysis

The section involved analysing the most frequent ASB types, ASB trends over time, and ASB distribution by Boroughs.

Analysing the most frequent ASB types

```
# Most frequent ASB types
plt.figure(figsize=(12,6))
sns.countplot(y=df['Opening_Type_1'], order=df['Opening_Type_1'].value_counts().index, palette="coolwarm")
plt.title("Most Frequent ASB Types")
plt.xlabel("Count")
plt.ylabel("ASB Type")
plt.show()
```

Figure 12: Code for analysing the most frequent ASB types

After running the code, the following output came up:

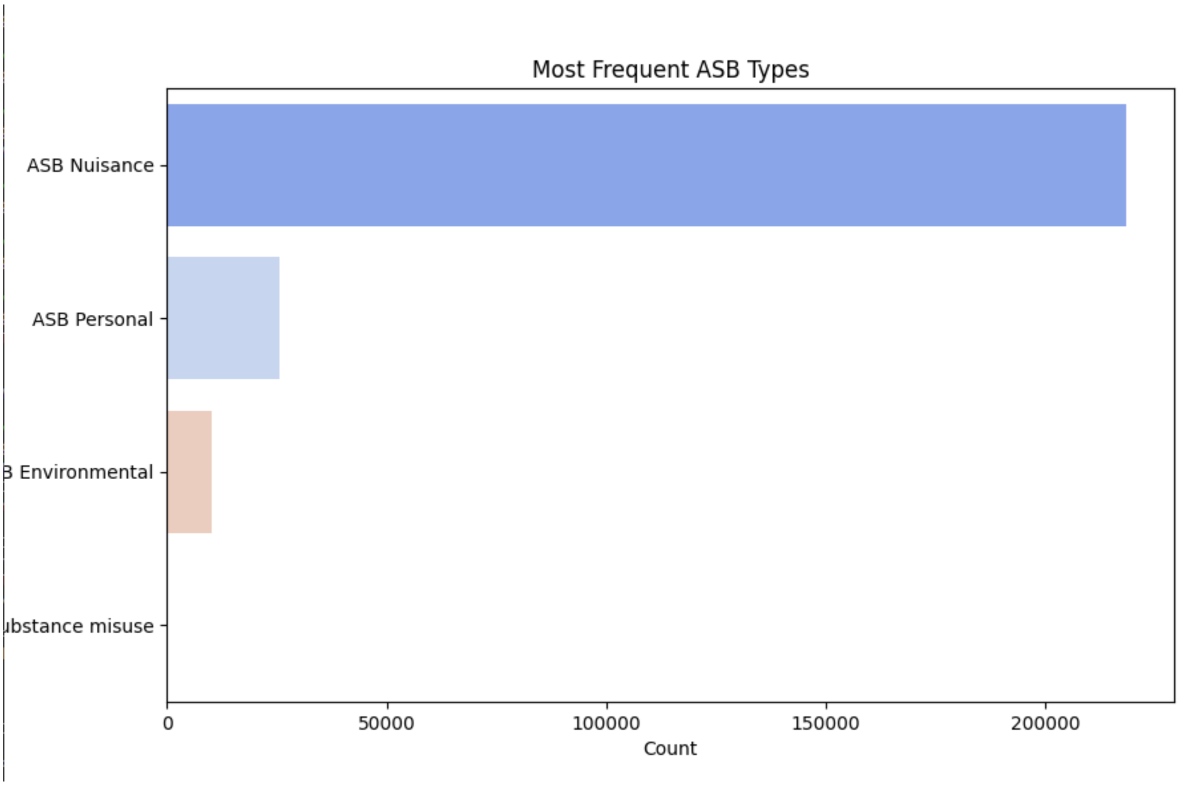

Figure 13: The most frequent ASB types

According to the results ASB nuisance was most frequent ASB type followed by ASB personal. Substance abuse was the last on the list. The law enforcement agents should therefore put extra effort in dealing with ASB nuisance because it is the most prevalent ASB type. Efforts should also be put in place to deal with the other ASB types.

Analysing the most ASB Trends Over Time

The code below was used to analyse the most ASB trends over time:

```
# Plot ASB incidents by hour
plt.figure(figsize=(12,6))
sns.histplot(df['Hour'], bins=24, kde=True, color="blue")
plt.title("ASB Incidents by Hour")
plt.xlabel("Hour of the Day")
plt.ylabel("Incident Count")
plt.show()
```

Figure 14: Plot ASB incidents by hour

Running the code led to the output in the screenshot below:

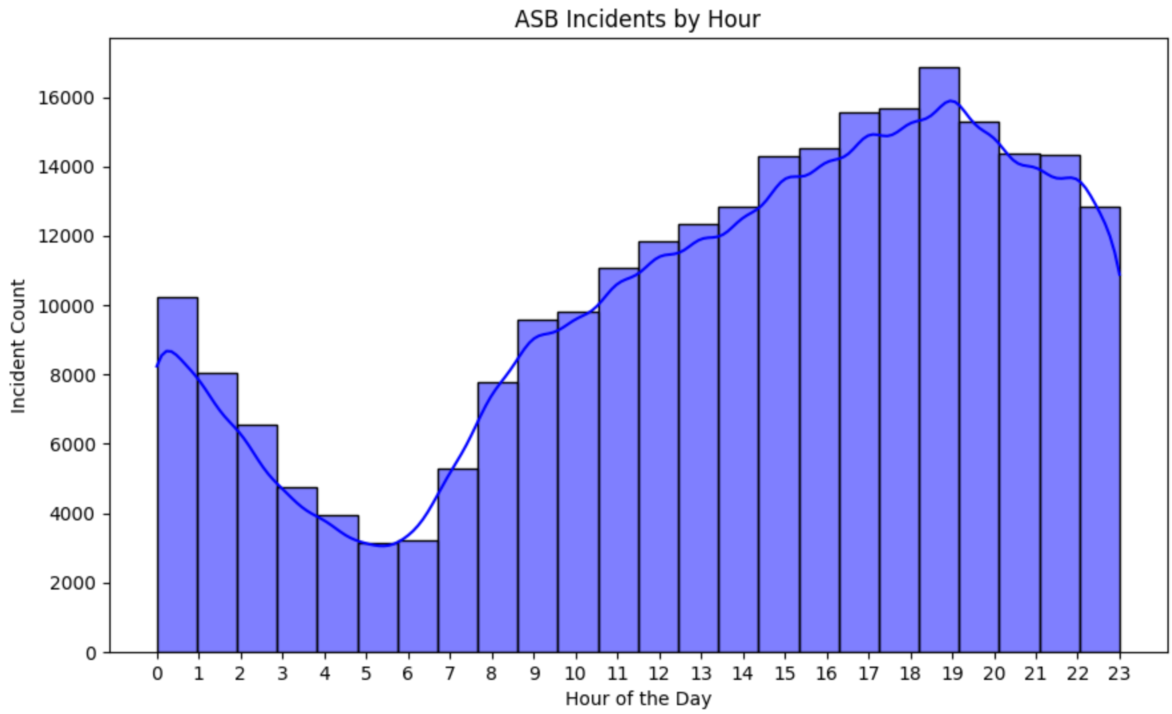

Figure 15: ASB incidents by hour

Each bar represents a different hour from midnight (0) to 11 PM (23). Incidents frequently occurred between 18:00 and 19:00. The least incidents in the day occurred between 5:00 and 7:00. The information is important for the policing department because it will help in scheduling patrols or putting in place the necessary measures that will help in dealing with the incidences.

ASB distribution by Boroughs

The code below was used to visualize ASB distribution by boroughs:

```
# ✅ New Visualization: ASB Incidents by Borough
if 'Safer_Neighborhood_Team_Borough_Name' in df.columns:
    plt.figure(figsize=(14,7))
    sns.countplot(y=df['Safer_Neighborhood_Team_Borough_Name'], order=df['Safer_Neighborhood_Team_Borough_Name'].value_counts().index, palette=
"viridis")
    plt.title("ASB Incidents by Borough")
    plt.xlabel("Count")
    plt.ylabel("Borough")
    plt.show()
else:
    print("⚠ 'Safer_Neighborhood_Team_Borough_Name' column not found. Skipping borough-wise ASB visualization.")
```

Figure 16: Visualizing ASB distribution by boroughs

The output of the code is shown below:

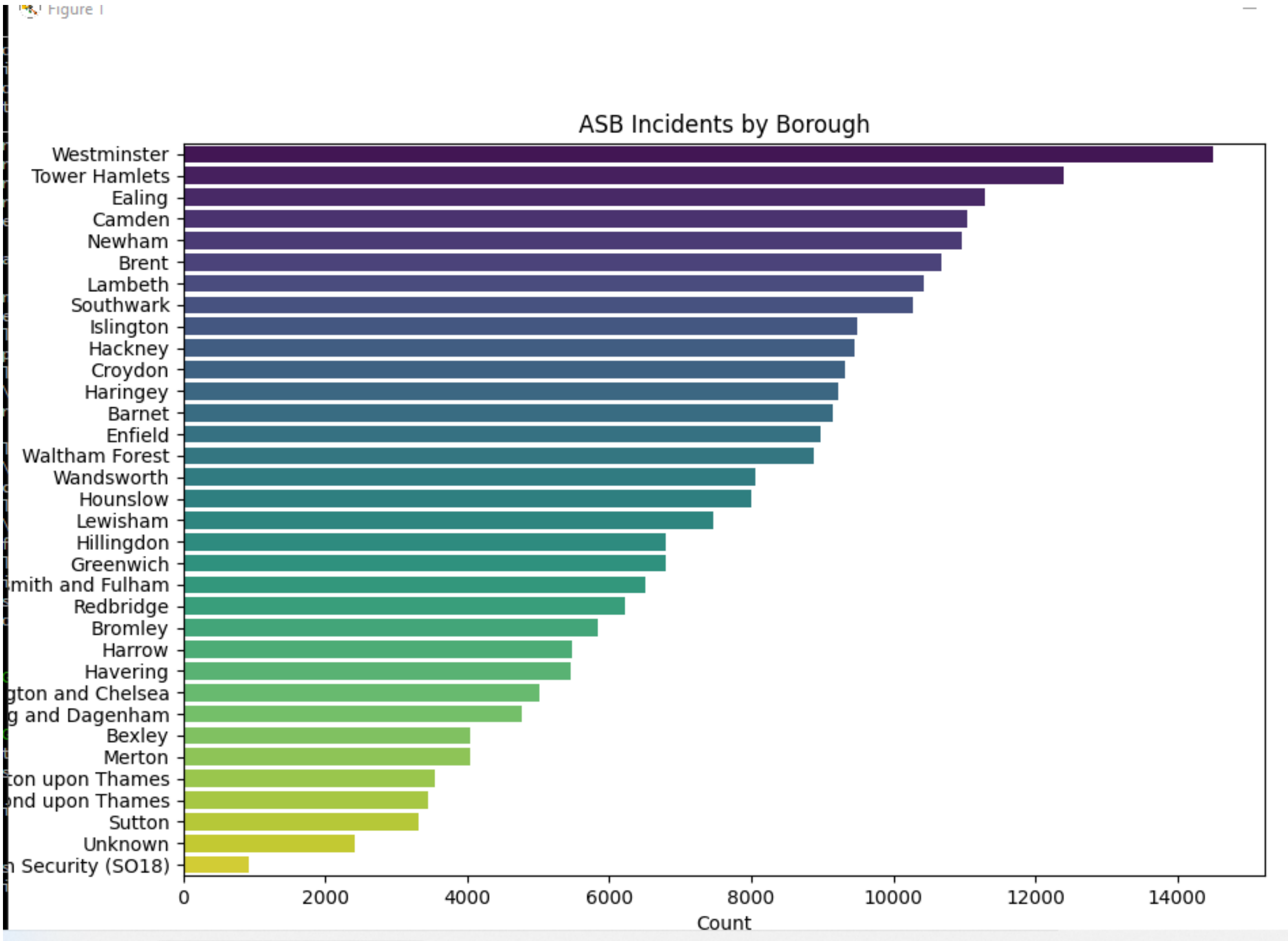

Figure 17: ASB incidents by borough

Most of the incidents were recorded in Westminster whereas the least incidents were recorded in S)18. The information is important for the law enforcement agencies because it enables them to focus their efforts in the areas that have more incidents and require more attention as opposed to the areas that do not require so much attention.

The steps taken to solve the problem included data cleaning and processing that handled the missing values, eliminated the duplicates then formatted timestamps, exploratory data analysis that visualized ASB patterns across boroughs, time and types as well as Hypothesis testing using Chi-Square and T-tests to validate key assumptions.

Statistical Hypothesis Testing

Chi-Square Test for ASB Incidents Across Boroughs

Hypothesis:

- **Null Hypothesis (H_0):** The distribution of ASB incidents is uniform across all boroughs.
- **Alternative Hypothesis (H_1):** Certain boroughs have significantly higher ASB incidents than others.

The steps to test the hypothesis included creating contingency table of ASB incidents per borough then apply Chi-Square test to check for significant differences. The code below was used for carrying out the Chi-square testing:

```
# --- Hypothesis Testing ---

# ✅ 1️⃣ Chi-Square Test for ASB Incidents Across Boroughs
print("\n📊 Performing Chi-Square Test for ASB Distribution Across Boroughs...\n")
if 'Safer_Neighborhood_Team_Borough_Name' in df.columns:
    # Create a frequency table of ASB incidents per borough
    contingency_table = df['Safer_Neighborhood_Team_Borough_Name'].value_counts().to_frame()

    # Perform the Chi-Square Test
    chi2_stat, p_value, dof, expected = chi2_contingency(contingency_table)

    # Print results
    print(f"Chi-Square Statistic: {chi2_stat}")
    print(f"P-value: {p_value}")
    print(f"Degrees of Freedom: {dof}")

    # Decision Rule
    if p_value < 0.05:
        print("❌ Reject H₀: ASB incidents are not uniformly distributed across boroughs.")
    else:
        print("✅ Fail to Reject H₀: ASB incidents are uniformly distributed across boroughs.")
else:
    print("⚠️ 'Safer_Neighborhood_Team_Borough_Name' column not found. Chi-Square test skipped.")
```

Figure 18: Chi-Square for hypothesis testing

Running the code brought out the following output:

📊 Performing Chi-Square Test for ASB Distribution Across Boroughs...

Chi-Square Statistic: 0.0
P-value: 1.0
Degrees of Freedom: 0
☑ Fail to Reject H_0 : ASB incidents are uniformly distributed across boroughs.

Figure 19: Performing a Chi-Square test on the dataset

The output from the performance test for ASB (Anti-Social Behaviour) distribution across borough provides statistical results to determine whether the incidents are uniformly distributed. The Chi-Square statistic was 0.0 which suggests that there is no difference, which means that the observed data perfectly matches the expected uniform distribution. P-value of 1.0 helps in the identification of the significance of results. A p-value of 1.0 indicates that there is no significant evidence available that can make us to reject the null hypothesis (H_0). In the current case, there is no evidence that suggests that ASB incidents are not uniformly distributed. The degrees of freedom are zero, which is the number of values in the final calculation of a statistic that are free to vary. A value of 0 indicates that the data fits the expected distribution perfectly. We therefore fail to reject the null hypothesis on the basis of the chi-square statistic and the p-value. It therefore means that ASB incidents are considered to be uniformly distributed across the boroughs.

📊 Performing T-Test for ASB Incidents Between High and Low Boroughs...

T-Statistic: 11.672103650625862
P-value: 9.163972109200561e-10
✕ Reject H_0 : High ASB boroughs have significantly more incidents than low ASB boroughs.

Figure 20: Performing T-test on the data

The output of the T-test for ASB (Anti-Social behaviour) incidents between high and low boroughs provides statistical results to determine if there is a significant difference in the number of incidents between these two groups. The calculation gave a T-statistic of 11. 672103650625862, which indicates that the size of the difference relative to the variation in the sample data. A large T-statistic suggests a significant difference between the groups. The p-value was 9.163972109200561e-

10 which is extremely small (much less than 0.05), indicating strong evidence against null hypothesis (H_0). This suggests that the difference in ASB incidents between high and low boroughs is statistically significant. The conclusion is there to reject the alternative hypothesis that high ASB boroughs have significantly more incidents than low ASB boroughs.

Task 4: Result Evaluation and Future Development

The results from the T-test that indicate that high ASB boroughs have significantly more incidents than low ASB boroughs can have several important implications. First, it can help in resource allocation in such a way that authorities can prioritize resource allocation to high ASB boroughs, ensuring that these areas receive more attention in terms of policing, community programs and preventive measures. Second, policy makers can use the data to come up with targeted strategies that will reduce ASB in high incidence areas, which can lead to more effective use of public funds. Third, understanding the disparity can help in engaging local communities effectively, fostering community-led initiatives to tackle ASB.

The results are significant but there are several limitations that should be considered. First, the accuracy of the results will depend on the quality and completeness of the incident data. In the case where the data is biased or inaccurate, the validity of the findings could be affected. Second, the analysis fails to account for contextual factors such as socio-economic conditions, population density or local policies that are likely to influence ASB rates. Third, the analysis is a snapshot in time and does not consider seasonal or temporal variations in ASB incidents. Fourth, the definition of boroughs might not capture micro-level variations within boroughs, potentially masking localized hotspots.

Future work and enhancements should be done to ensure that the data is useful in real life situation. First, machine learning models should be incorporated to predict ASB hotspots and trends, which will allow for proactive rather than reactive measures. Second, real-time data processing and visualization tools should be implemented to provide up-to-date insights for immediate action. Additional datasets such as socioeconomic data, urban infrastructure should be integrated to provide more comprehensive understanding of the factors that contribute to ASB.

Section 2: Business Intelligence (Tableau) – Examining mental health impacts on remote employees using *Remote_Work_and_Mental_Health.csv*.

Task 1: Visualize 5 of the most impactful variables on mental health in descending order.

The first step was to identify key variables that affect mental health such as stress level, social isolation, job satisfaction, work-life balance, and mental health conditions. The next step was to create a bar chart that shows their impact in descending order on the basis of survey responses. I then sorted the values from highest to lowest for clarity.

How stress levels affect mental health

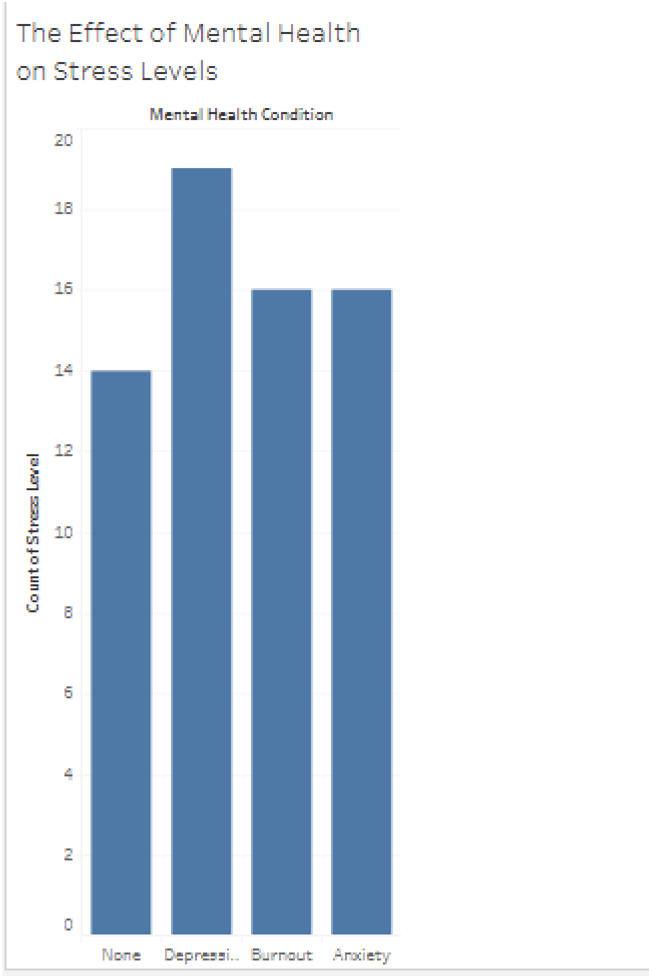

Figure 21: How stress levels affect mental health

People with anxiety and burnout recorded the highest stress levels which shows how the two variables correlate. The stress levels among people with depression were also high but lower than the stress levels among people with anxiety and burnout. People with no mental health conditions also recorded high stress levels, which is indicative of the fact that people are facing challenges that stress them in their lives or in their workplace but they have not yet developed mental health conditions.

How social isolation affects mental health

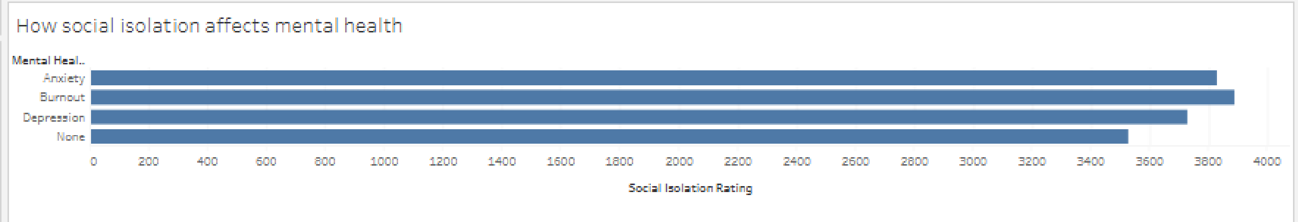

Figure 22: How social isolation affects mental health

Social isolation made more people to have burnout compared to the other mental health conditions. The rate at which people in social isolation experienced anxiety was also high but lower than the people with burnout. Depression was also recorded among the socially isolated persons at a high rate, but the rate was lower than the one recorded in the cases of anxiety and burnout. Additionally, a good number of the socially isolated individuals did not have any mental condition.

How job satisfaction affects mental health

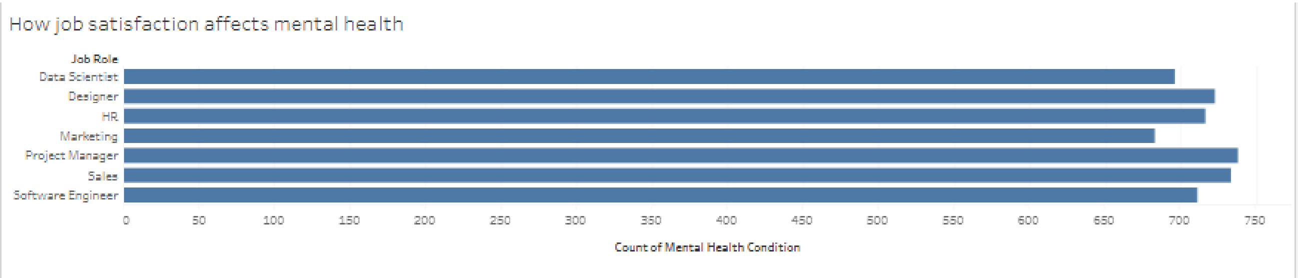

Figure 23: How job satisfaction affects mental health

Job titles are also associated with stress levels. For instance, the highest count of mental health conditions was recorded among the project managers and the people in sales, which is indicative of the fact that their jobs are very demanding. The lowest rate of mental health conditions was recorded among the people in the marketing field. It is important to note that the difference in the number of people with mental health conditions did not differ very much among the careers.

How work-life balance affects mental health

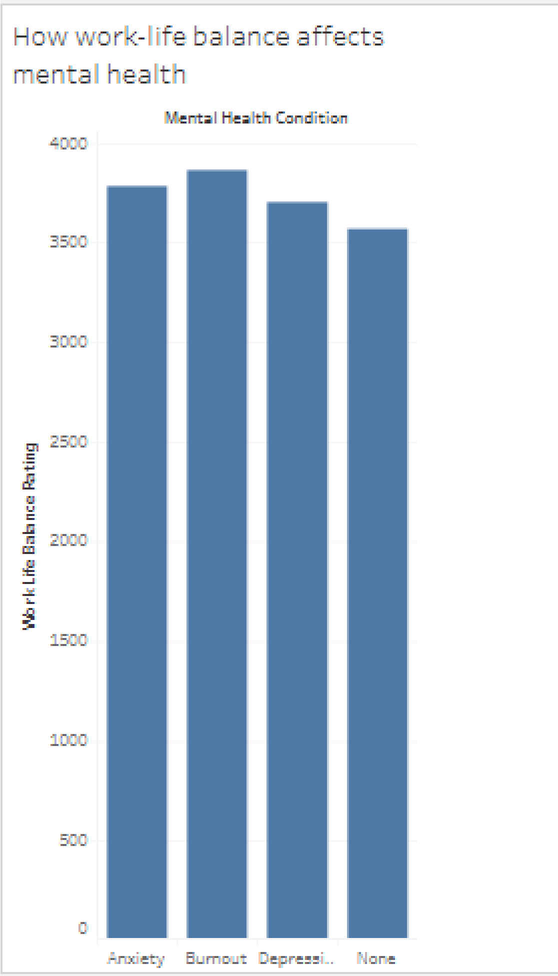

Figure 24: How work-life balance affects mental health

The highest work life balance was recorded among the people with burnout which is indicative of the fact that the people face many challenges in a bid to keep the work life balance. People with anxiety and depression also recoded a higher work life balance rating. Workers therefore do a lot to strike a work-life balance to the point that it affects their mental health.

How work location affects mental health

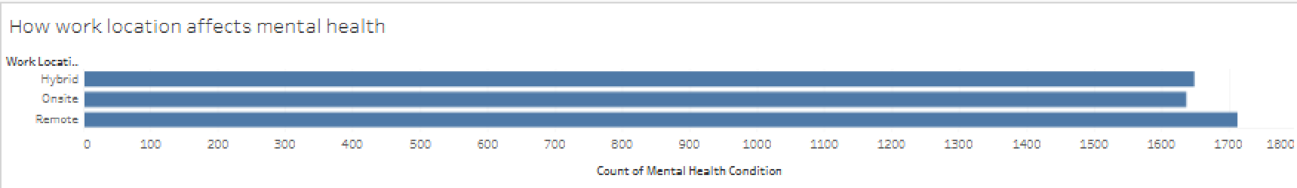

Figure 25: How work location affects mental health

The highest count of mental health conditions was recorded among the people that are working remote. Followed by the people that are working hybrid. It means that it is stressful to work in the two locations. It may be the case because of the loneliness that is associated with working remotely.

Task 2: Industry Vs stress Levels (Dependency Analysis)

The steps in task 2 included using a heatmap or grouped bar chart to analyse how different industries experience stress levels. The next step was to aggregate levels by industry then highlight the industries that have high stress.

How the industries affect stress levels

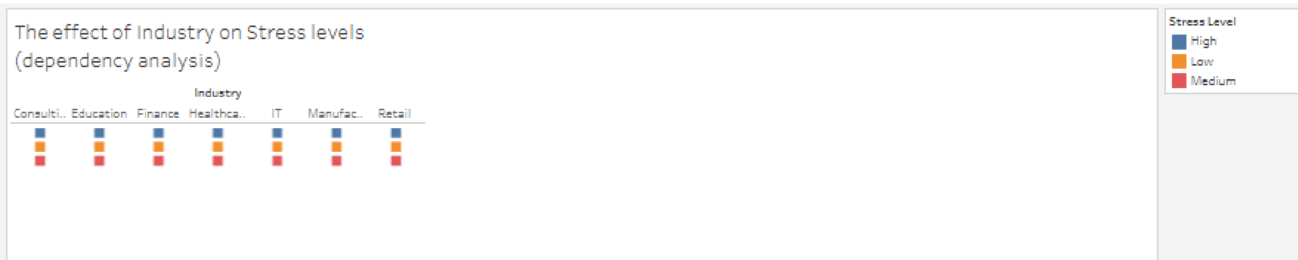

Figure 26: How the industries affect stress levels

Similar stress levels were recorded in all the fields which means that all the employees had an equal chance of suffering from mental health conditions regardless of their industry. Similar measures should therefore be put in place to deal with stress levels in different industries in the best way possible. However, it is important to note that the highest stress levels were recorded among the people that work in the finance, IT, healthcare and retail fields. On the other hand, low stress levels were recorded among the people working in the education field which is indicative of the fact that working in the education industry is less costly compared to working in the other fields.

Task 3: Type of Work (Onsite, Hybrid, Remote) vs. Stress Levels

The steps in the current task include comparing stress levels for each work type (Onsite, hybrid, remote) then using a box plot to analyse the stress distribution across the work types.

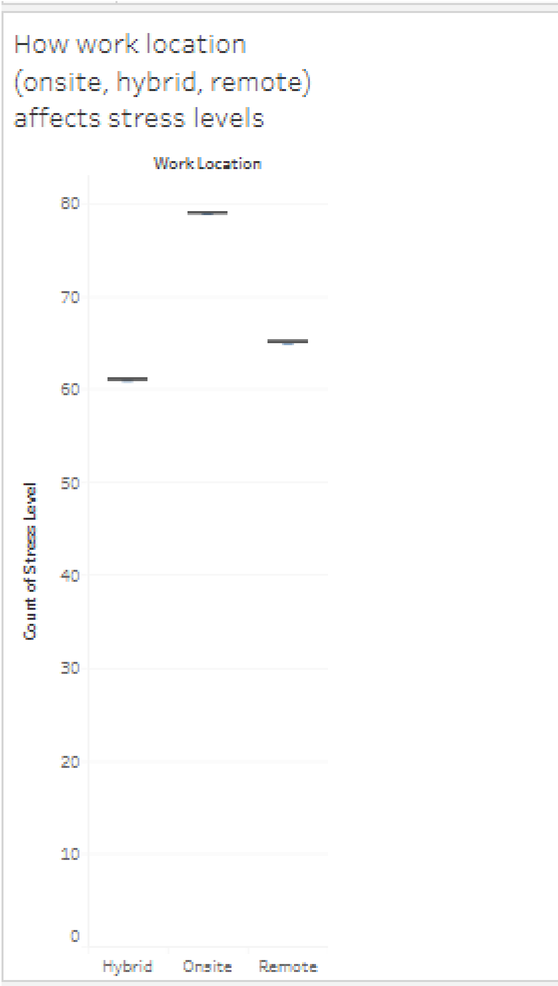

Figure 27: How work location affects mental health

The highest count of mental health conditions was recorded among the people that are working remote. Followed by the people that are working hybrid. It means that it is stressful to work in the two locations. It may be the case because of the loneliness that is associated with working remotely.

Task 4: Work-Life Balance vs. Mental Health Conditions

The task involved determining how work-life balance affects mental health. A bar chart was used to compare the satisfaction levels and reported conditions.

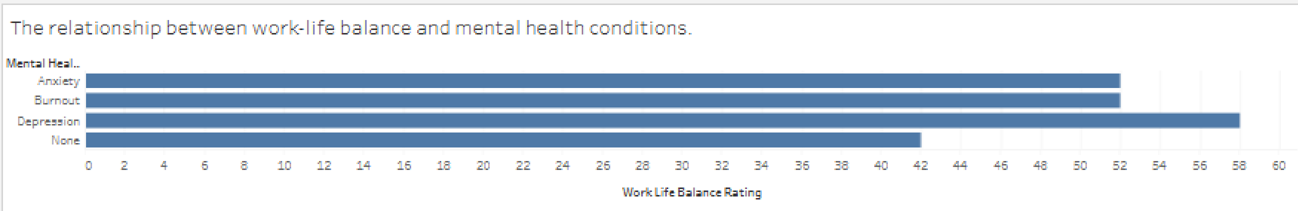

Figure 28: Work-Life Balance vs. Mental Health Conditions

The highest work life balance was recorded among the people with burnout which is indicative of the fact that the people face many challenges in a bid to keep the work life balance. People with anxiety and depression also recoded a higher work life balance rating. Workers therefore do a lot to strike a work-life balance to the point that it affects their mental health.

Task 5: Create an Interactive Dashboard

The task involved creating an interactive dashboard. I combined at least four sheets into a single dashboard. I then added filters then ensured interactivity by clicking one sheet updates all others.

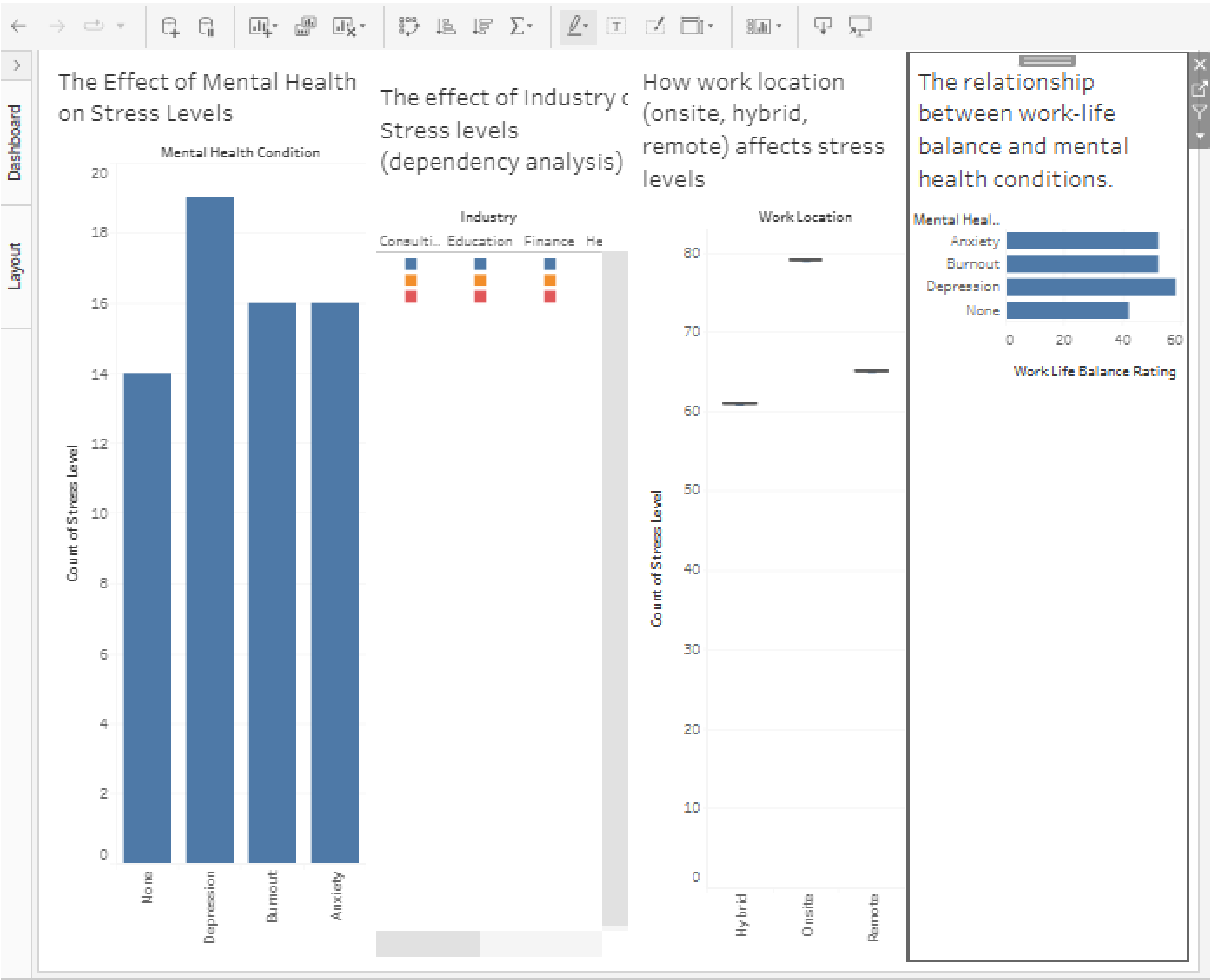

Figure 29: Dashboard created

References

Esposito, C., Affuso, G., Miranda, M. C., & Bacchini, D. (2021). A new dimensional measure of antisocial behaviour evaluation (ASBE). *European Journal of Developmental Psychology*, *18*(2), 257-270.

Murray, S. (2022). *Spatiotemporal Crime Prediction in Northern Ireland* (Doctoral dissertation, Ulster University).

Schroeders, U., Mariss, A., Sauter, J., & Jankowsky, K. (2023). Predicting Juvenile Delinquency and Criminal Behavior in Adulthood Using Machine Learning.

Singh, R., Subramani, S., Du, J., Zhang, Y., Wang, H., Miao, Y., & Ahmed, K. (2023). Antisocial behavior identification from twitter feeds using traditional machine learning algorithms and deep learning. *ICST Transactions on Scalable Information Systems*, *10*(4).

Viding, E., McCrory, E., Baskin-Sommers, A., De Brito, S., & Frick, P. (2024). An ‘embedded brain’approach to understanding antisocial behaviour. *Trends in Cognitive Sciences*, *28*(2), 159-171.

Appendix

Goggle Collab link:

https://drive.google.com/file/d/12J5XEMnXyPSm5KFStyhbVvlpvg3HNOoX/view?usp=sharing

Tableau link:

https://public.tableau.com/app/profile/walter.zenia/viz/Remote_Work_and_Mental_Health/Task2IndustryVsstressLevelsDependencyAnalysis?publish=yes