**The British College**

**KATHMANDU**

**Coursework Submission Coversheet**(individual coursework only)

**Faculty of Arts, Environment and Technology LBU Student Id:**

77261075

**For checking by the student:**

Please ensure all information is complete and correct and attach this form securely to the front of your work before posting it in a coursework collection box.

Award name: Bsc Hons in Computing

Module code: [15781 - AUT - 202210](https://my.leedsbeckett.ac.uk/webapps/blackboard/execute/courseMain?course_id=_157454_1)

Module name: Advance Database System (ADS)

Module run: 2022

Coursework title: Data warehouse Design and Development

Due Date:

Module leader: (In LBU): Jackie Campbell, Sanela Lazarevski

Module tutor: (In TBC): Dibya Tara Shakya

**TURNITIN** Checked: YES NO ***(please circle)***

Submission date& time: Date: Time: Before noon

**Total Word Count: Total Number of Pages (including this front sheet):**

**In submitting this form with your assignment, you make the following declaration:**  
I declare, that the coursework submitted is my own work and has not (either in whole or part) been submitted towards the award of any other qualification either at LBU or elsewhere. I have fully attributed/referenced all sources of information used during the completion of my assignment, and I am aware that failure to do so constitutes an assessment offence.

Signed: Mavira Bhattarai Date: <Submitted Date>

**You are strongly advised to retain a second copy of your work in case of any query about the assignment.**

**For completion by the faculty:**

**This mark is provisional and subject to moderation and approval by the relevant examining board**

**Teacher's Feedback**

**Teacher's Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Contents

[Assignment 2 – Tasks 1 4](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432923)

[**Task 1: Data Analysis/OLAP/Mining Investigation** 4](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432924)

[Part A: Spread sheet Reports 4](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432925)

[Generating Charts and Reports: 8](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432926)

[MS-Excel Pivot Tables, Reports and Charts 8](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432927)

[Part B: Data Analytics and Dashboards 14](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432928)

[Apex Dashboard Reports and Charts 15](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432929)

[Tableau Reports and Charts 26](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432930)

[Bibliography 43](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432937)

[Appendix 44](file:///C:\Users\Dell\AppData\Local\Temp\Temp1_ADSII_AshimKumarShrestha_1.18.2022.zip\ADSII_AshimKumarShrestha_1.18.2022\Component%202%20Word%20Document.docx#_Toc93432938)

# Assignment 2 – Tasks 1

## Task 1: Data Analysis/OLAP/Mining Investigation

### Part A: Spread sheet Reports

**Data** analysis is the process of establishing a set of data, looking into it, cleaning it up, and converting it to produce actionable business decisions and results. Data analysis is regarded as a subfield of data analytics, which is used in businesses to examine data and derive certain insights from it in order to make wise judgments. The process of identifying errors and developing measures to prevent them from happening again in the future is one that takes into account the influences of specific actions that were made at the time. Data Analysis consists of the following processes:

* Data Requirement Gathering
* Data Collection
* Data Cleaning
* Data Analysis
* Data Interpretation
* Data Visualization

**Business intelligence (BI)** is a term used to describe a collection of methods or technology-driven procedures that help businesses or organizations evaluate/analyze data and business information, improve decision-making abilities, carry out data mining tasks, and provide business users with useful data visualizations for strategic planning. By combining analytics, data management, and reporting tools with various additional techniques for the management and analysis of data, BI aids in these sectors. Business intelligence is primarily used to assist organizations in expanding their businesses by enhancing decision-making processes at all levels of management as well as tactical and strategic management processes**.** The key benefits of having BI integrated in a business are mentioned below:

* Enhance and accelerate decision-making
* Enhance internal business procedures.
* Increase productivity and operational effectiveness
* Recognize issues with your company that need to be fixed
* Recognize new market and business trends.
* Create more effective business plans
* Promote increased sales and new revenue
* Gain an advantage over competing businesses

Data analysis is a step in the business intelligence process, and one of the most helpful tools for interpreting the findings of business intelligence and data analysis is data visualization. The graphical display of information and data using visual components like charts, graphs, and maps is known as data visualization. In the modern world, it is absolutely necessary to analyze the huge amounts of data and make data-driven decisions, and data visualization tools offer an easy approach to examine and comprehend trends and patterns in data. Common general types of data visualization tools are:

* Charts
* Tables
* Graphs
* Maps
* Info graphics
* Dashboards

The following image shows how data is visualized using data visualization tools.

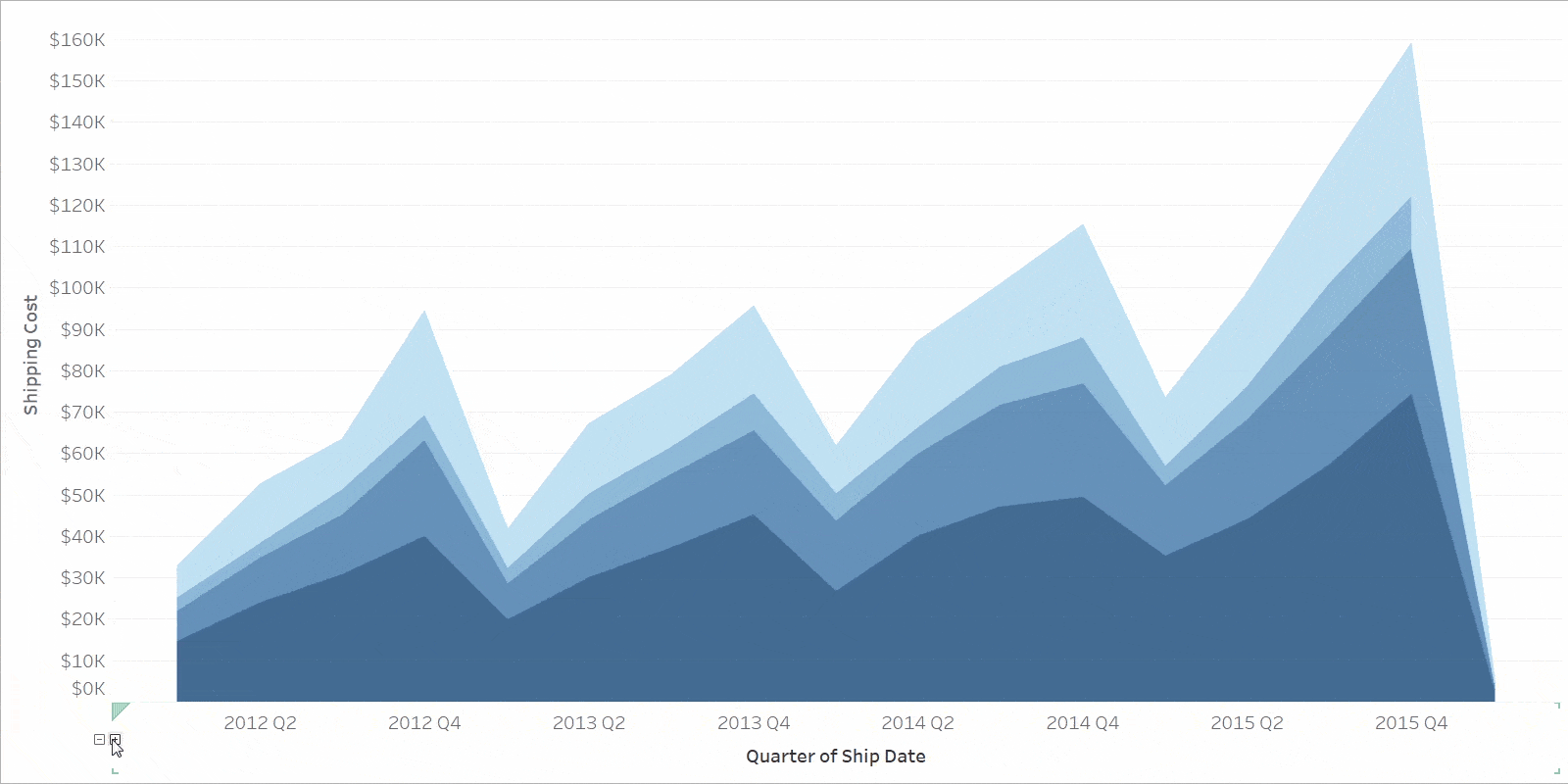


Figure : Example of data visualization using unique charts

The benefits of using data visualization tools are mentioned below:

* A more innovative and useful approach to comprehending information concealed in data
* A simpler method to discover the ideas we want
* A more effective and quick way to spot patterns, trends, and correlations in data sets that text- or figure-based reports would miss
* Increased likelihood of discovering new business possibilities when they present themselves as a result of forecasting future trends, sales volume, or revenue generation.

**Data mining** is one of the key subjects covered in data analytics. Data mining is the technique that businesses utilize to transform unusable, unprocessed data into useful information. The technique of looking for patterns while analyzing a huge amount of data enables businesses to better understand trends and create marketing plans that will ultimately result in higher sales and lower costs.

**Data Mining or KDD** (Knowledge Discovery of Data) has the main goal of finding patterns that helps in making predictions or insights using large data sets. It has a huge range of use like looking for patterns in the markets by financial sectors to governments trying to identify a potential security threat. Some key information to remember about data mining is as follows:

* Analyzing a vast amount of data to find patterns and trends is known as data mining.
* Data mining can be used by corporations for everything from learning about what customers are interested in or want to buy to fraud detection and spam filtering.
* Data mining programs break down patterns and connections in data based on what information users request or provide.
* Social media companies use data mining techniques to materialize their users in order to generate profit.
* This use of data mining has come under criticism lately as users are often unaware of the data mining happening with their personal information, especially when it is used to influence preferences. (TWIN, 2021)

## 

**OLAP (Online Analytical Processing)** is an online processing system that performs rapid multi-dimensional (categories) analysis of data over a huge volume of data with complex queries. It is used to deal with data warehouses, data marts or any other centralized data store and contains historical data of the organization that is later used for generating business insights. It retrieves data from several relational data sets in de-normalized form and sorts them into multiple dimensions that allow fast processing and analysis of data. ACID (Atomic, Consistent, Isolated and Durable) compliance is less applied on the data because OLAP focuses more on reading than modifying the data.

Features of OLAP:

* OLAP system uses a 3d-cube based multidimensional database that makes processing and analysis of complex data queries much easier and quicker than a traditional relational database.
* It also provides a drill-down approach for data analysis which leads to more granularity of a data in the database.
* The de-normalization of data in this system helps to improve the query processing speed significantly and allows few concurrent users at a time.

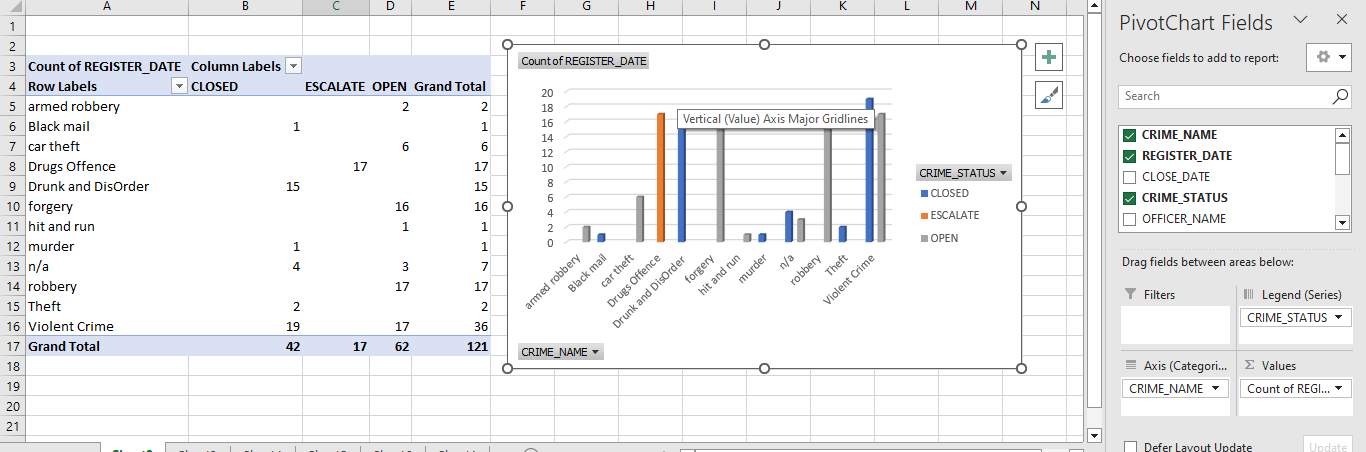
## Generating Charts and Reports:

The charts and reports required for this component are included below in the document and the reports have been generated using three different technologies; **Oracle Apex**, **Tableau** and **MS-Excel**. These technologies have been used to properly visualize and analyze data, and in addition to this, many drill-down reports that provide even more in-depth study of the data trend have also been built utilizing each technology.

### MS-Excel Pivot Tables, Reports and Charts

The reports that are listed below were produced using **MS-Excel** and are based on the information that was given to us so that we could complete this work. The necessary pivoting was done once the tables were downloaded, loaded, and loaded into the program.

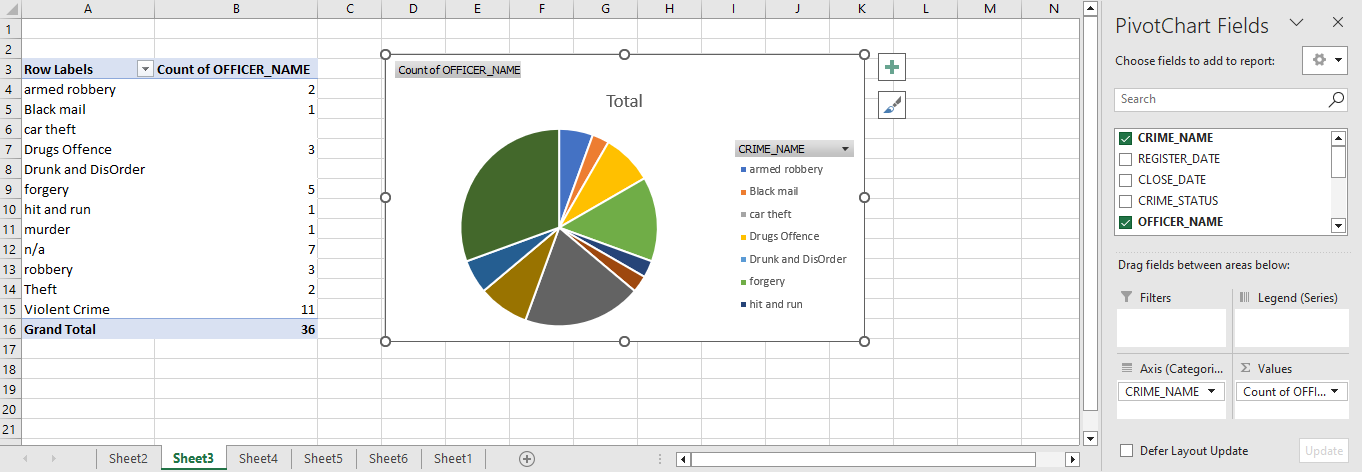
#### Report 1: Number of crimes per crime type along with crime status



The column chart given above shows the number of crimes per crime type with respective to the status of crime. The table with data was pivoted and the chart was created with proper display of data. The query that was used to generate this chart is also attached in the appendix of this document.

The trend in the chart shows that the number of crimes categorized as “Violent Crime” is the highest with a total occurrence value of 36. Similarly, the crime category with the highest escalated crime status is “Drugs Offence” with a total occurrence value of 17. The crime type with the highest number of closed crime status is violent crime i.e. 19 and with highest number of open crime status i.e. 17 are Robbery and Violent crime drugs and theft crimes.

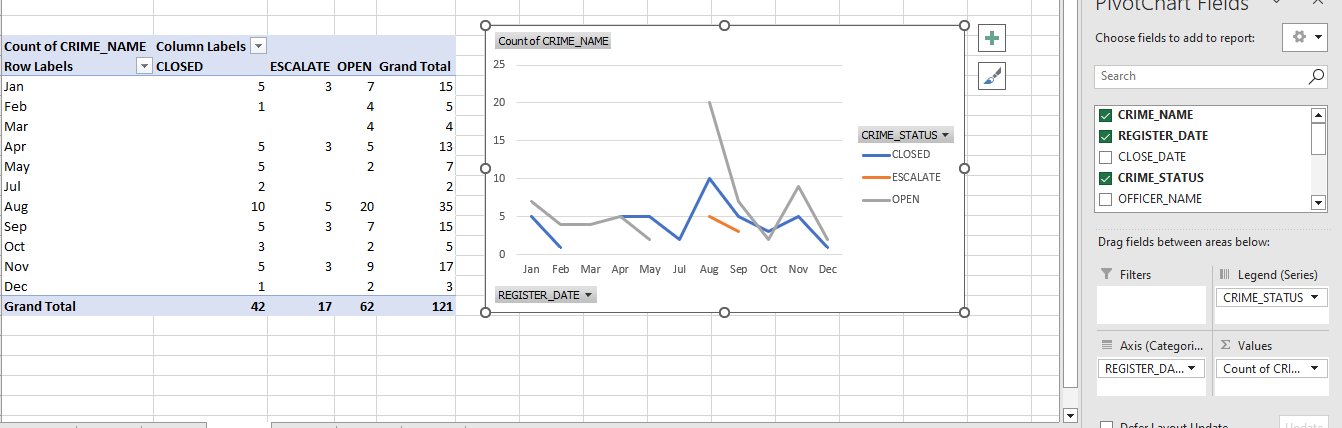
#### Report 2: Total number of officers per crime type



The doughnut chart displayed above shows the total number of officers per crime type.

We can see that the crime type “Violent Crime” has got the highest number of officers i.e. 11 whereas “Black mail”, “Hit and Run” and “Murder” have lowest number of officer i.e. 1.

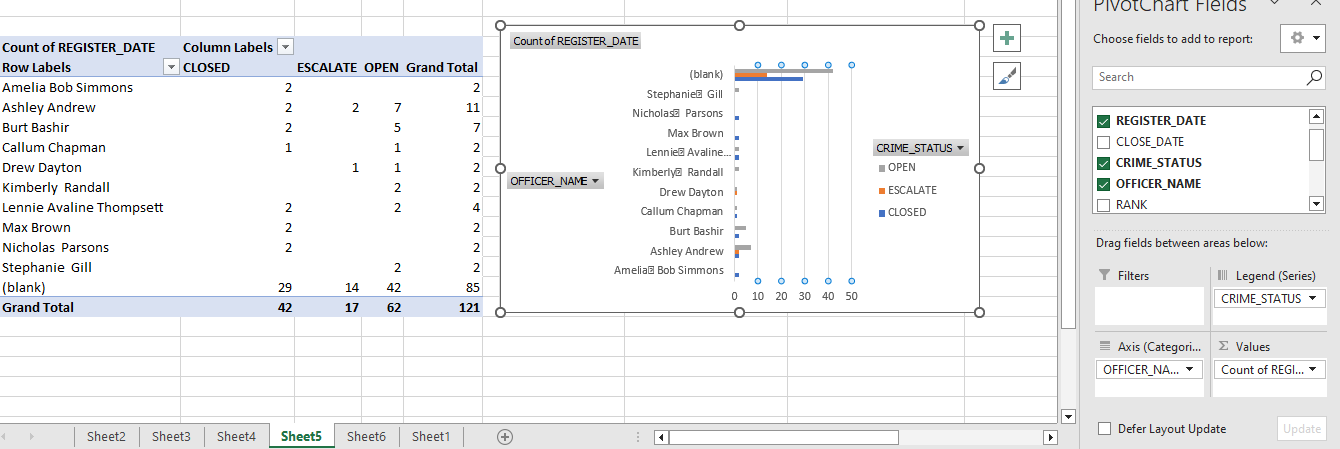
#### Report 3: Total number of crimes for all crime status per month



The column chart given above shows the total number of crimes with their crime status for each month the data source had the data about. The chart comprises the name of the month where the crimes were reported.

Talking about the trend in the chart, the chart clearly shows that the crime has closed, escalated and opened the most in August.

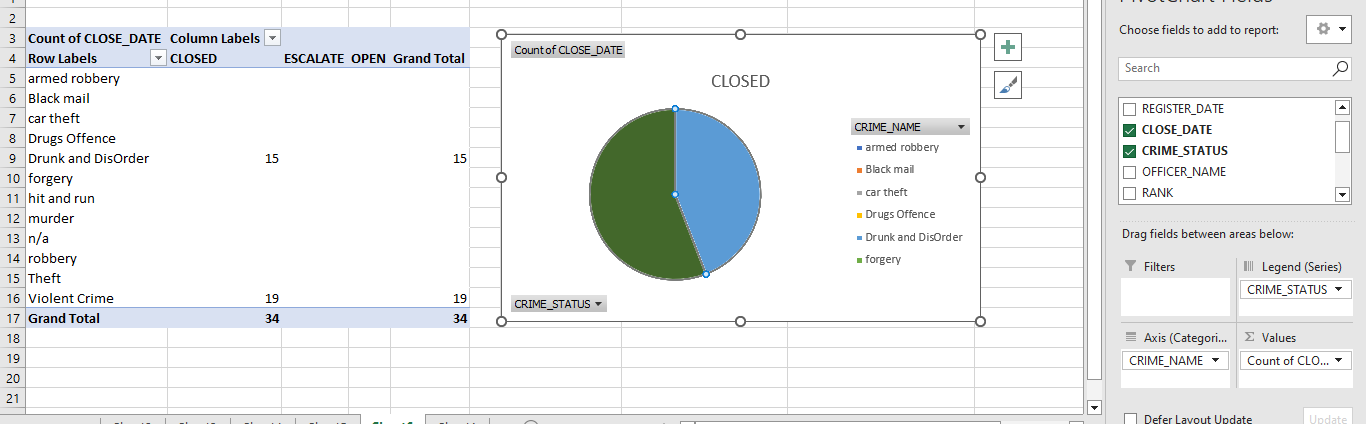
#### Report 4: Total number of crimes handled by each officer along with the crime status



The area chart displayed above holds the data for number of crimes handled by each officer along with the crime sttaus.

From the analysis of the data in the chart above, it was found that the office who has handled most of the cases was not mentioned in the data. However, in comparison to the mentioned officers, “Ashley Andrew” has handled the cases the most where 2 cases are closed, 2 cases are escalated and 7 cases are still open.

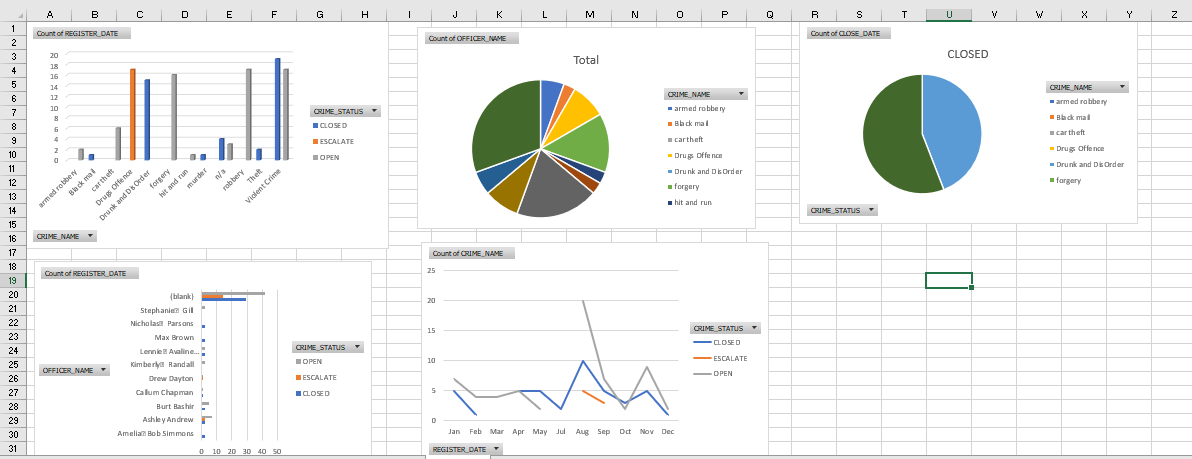
#### Report 5: Number of closed crimes per crime type



The pie chart given above total number of crimes that are closed per crime type.

The analysis of the data was done using the chart and it was found that only two crime types “Drunk and Disorder” and “Violent Crime” are closed with occurrence of 15 and 19 respectively.

Excel Dashboard



### Part B: Data Analytics and Dashboards

#### Data Analytics

Data analytics is the process of transforming unstructured or raw data into a format that is useful. The converted data helps with decision-making, drawing inferences, and implementing predictive analytical procedures after being cleansed, processed, and modeled. It is crucial for organizations because it aids in the optimization of their performance by lowering costs through the use of more effective methods to carry out business choices in the presence of a significant amount of data. Making sense of the data that is already available, analyzing it, and sharing greater business possibilities for the near future are the key goals of data analytics. Data analytics will provide an organization the benefits of:

* Identifying new opportunities and acquiring fresh insights to help them manage their business more effectively
* Providing meaningful information for making better business decisions.
* Enabling enterprises to make faster decisions, saving time and energy.
* Improving the strategies by focusing on the customer and improving the profits of the company.

The benefits provided by data analytics are similar to those provided by business intelligence; yet, they are used to generate different results. The primary purpose of data analytics is to model, cleanse, predict and transform the data as per the business needs. On the other hand, the primary purpose of business intelligence is to aid organizations to grow their businesses by providing support in decision making.

#### Dashboard

By gathering and gleaning valuable data from all the gathered data, a dashboard offers real-time data visualization. A dashboard compiles all the pertinent data that directly affects a company's success. A good dashboard has the advantages like:

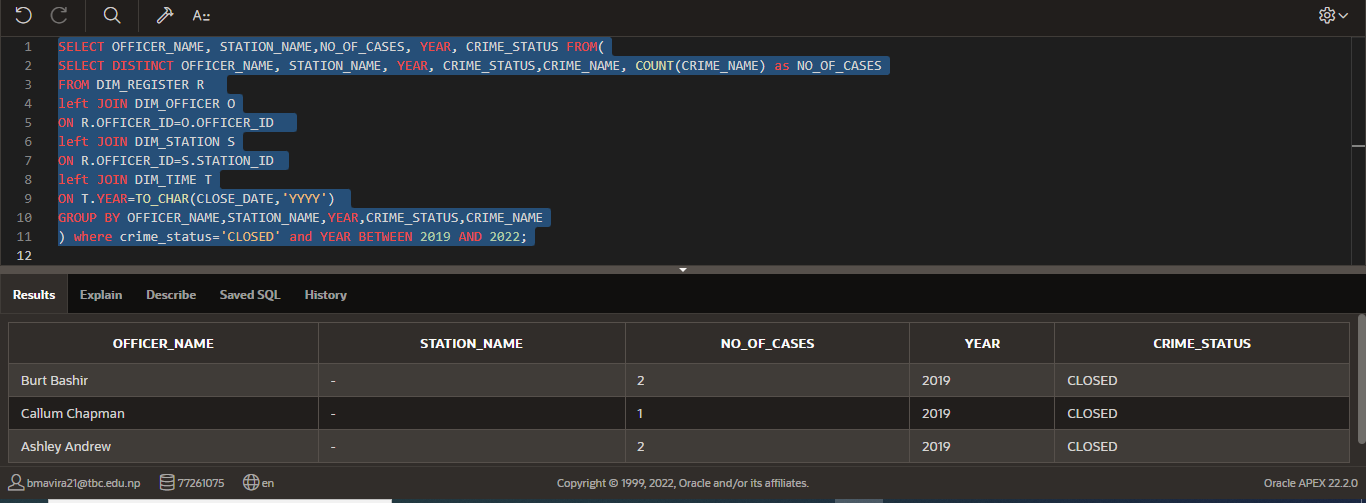
* Assisting in managing the complexity of running an organization by providing a full view of the important items that affect the organization.
* Consolidating important information from different departments and divisions within the organization
* Accessing critical information instantly and easily
* Displaying real time key metrics, performance indicators and assisting in making sound decisions
* Providing an easy way to navigate around the application.

Dashboards are a wonderful way to monitor your business, to look for answers, and to see all of your most-important metrics at a glance. The visualizations on a dashboard may come from one underlying dataset or many, and from one underlying report or many. A dashboard can combine on-premises and cloud data, providing a consolidated view regardless of where the data lives.

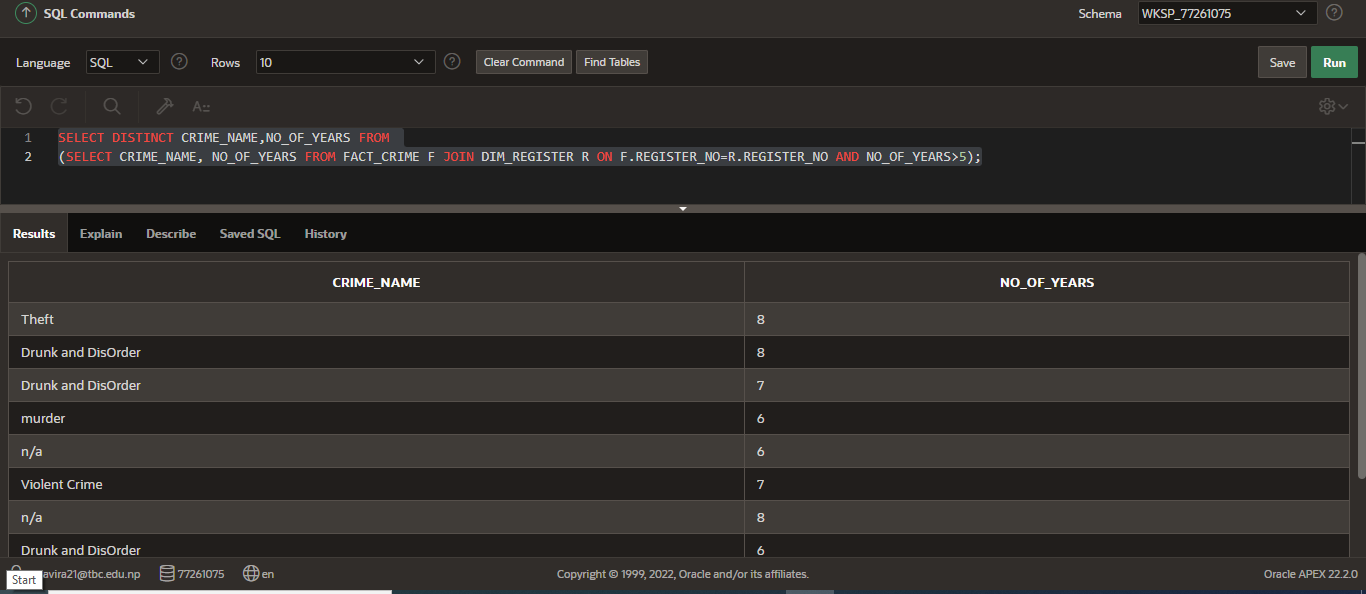
### Apex Dashboard Reports and Charts

The reports provided below were generated using **Oracle Apex** and are based on the data that were provided to us to be included in this task. The Leeds excel sheet have been utilized mostly for the creation of the charts below.

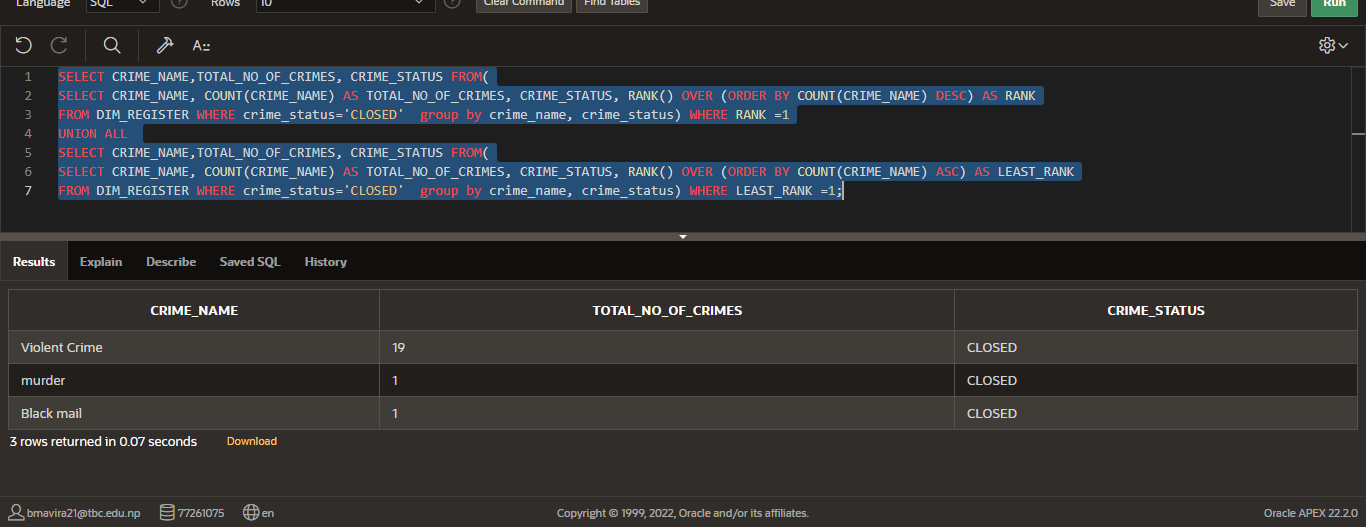
#### Report 1: Name of the police officer, number of crimes AND station name where the crime has closed from 2019 to 2022



#### Report 2: Crime name, time duration where time duration between registering crime and closing crime of the solved crimes is more than 2 years

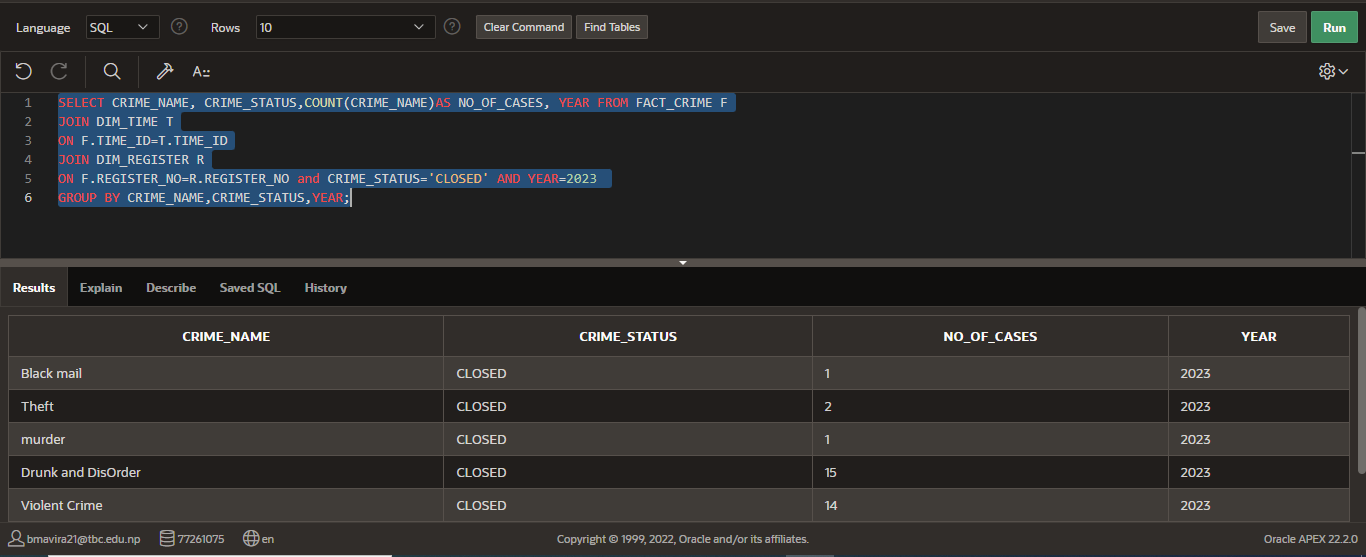


#### Report 3: Number of crimes and crime that has been closed the most and least

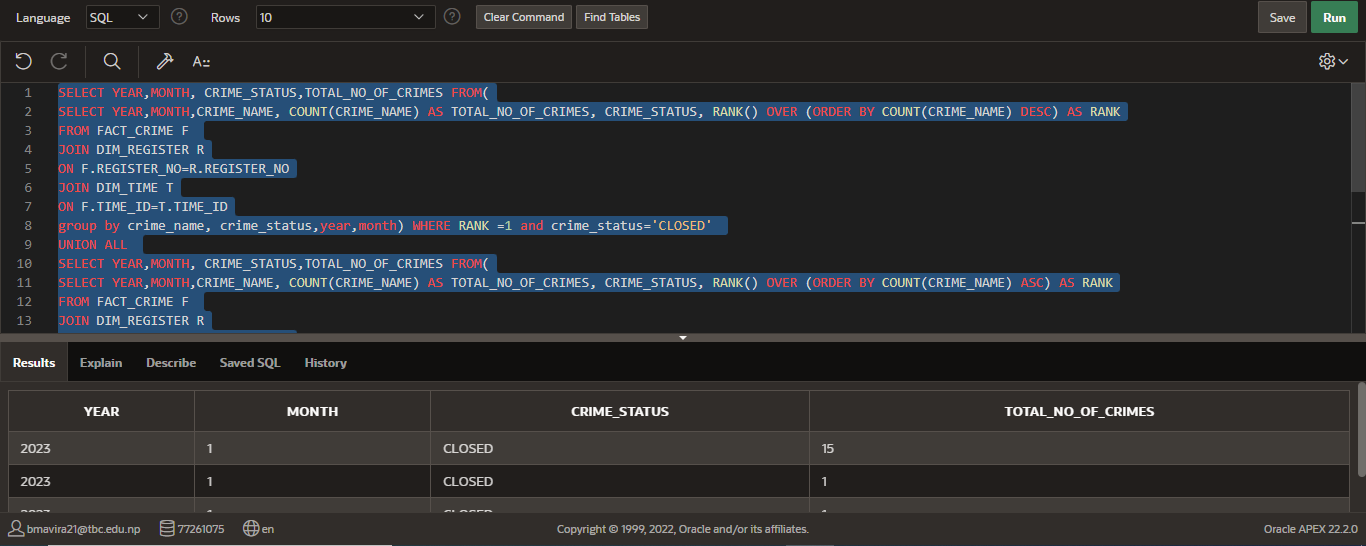


.

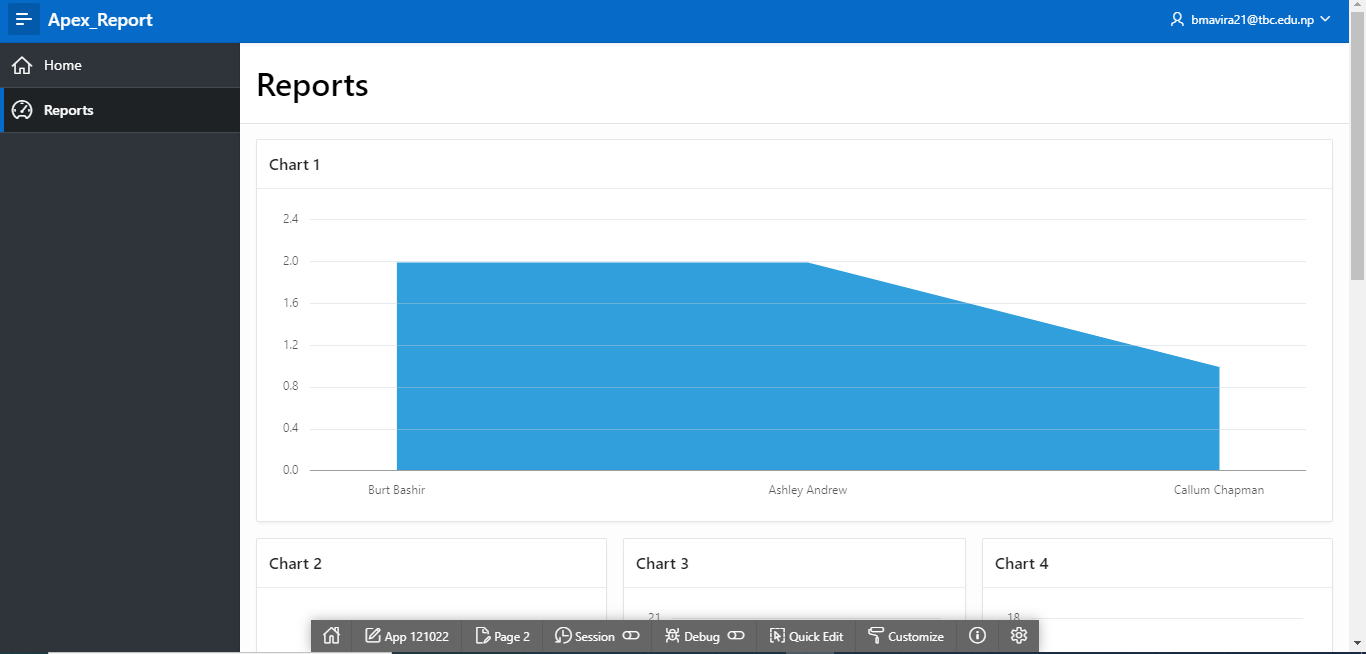
#### Report 4: Name of crime with total number that has been closed this year



#### Report 5: Year and month along with total number of crimes where crime has been closed the most and the least



Apex Dashboard



The chart 1 shows name of the police officer, number of crimes AND station name where the crime has closed from 2019 to 2022.

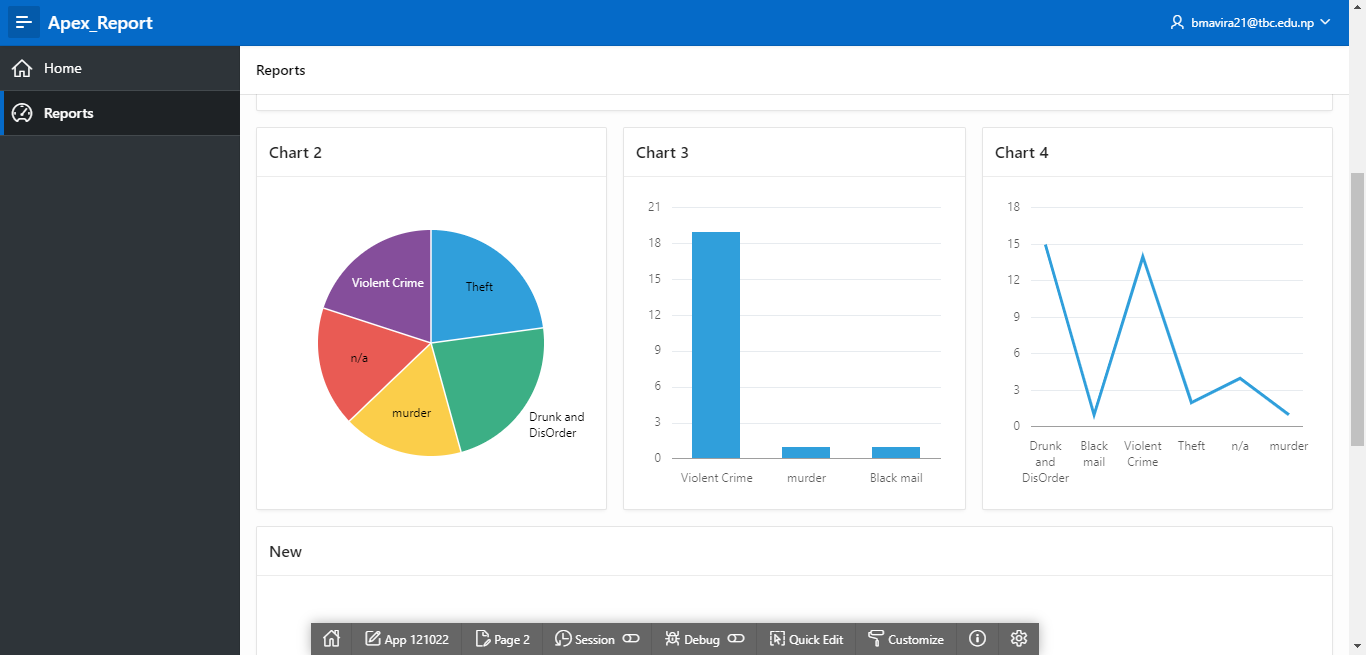


Chart 2 shows crime name, time duration where time duration between registering crime and closing crime of the solved crimes is more than 2 years.

Chart 3 shows number of crimes and crime that has been closed the most and least

Chart 4 shows name of crime with total number that has been closed this year

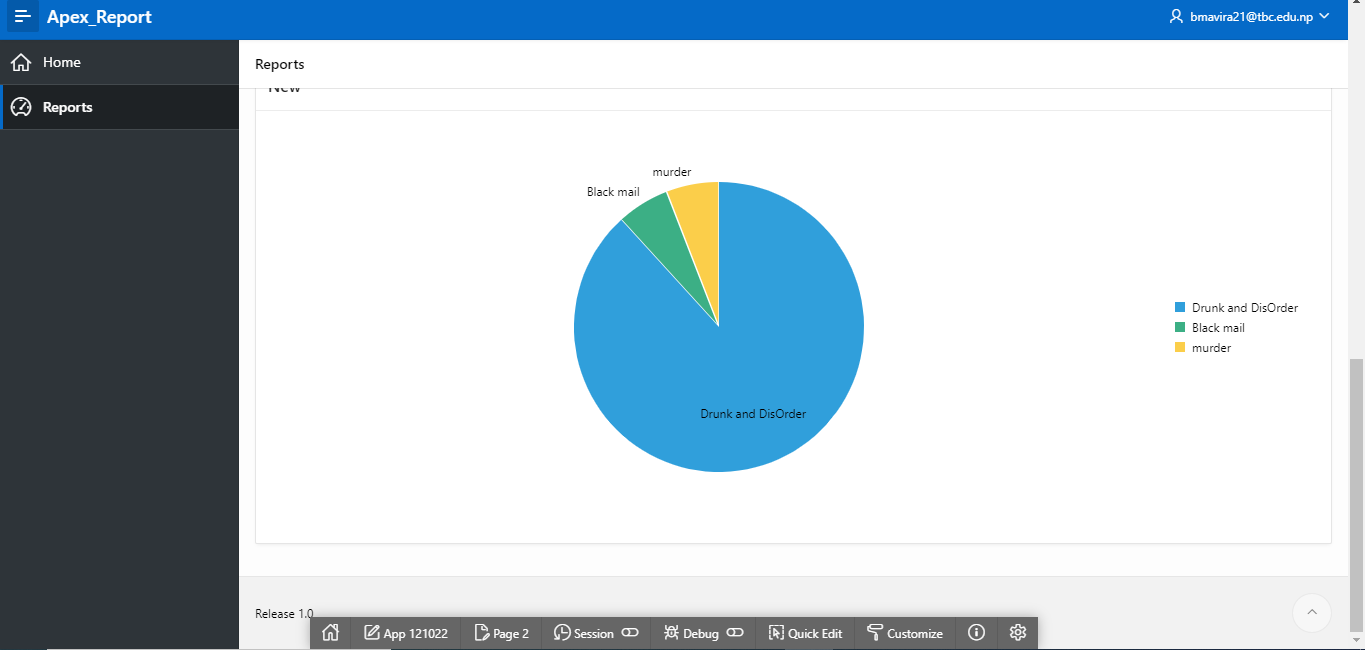
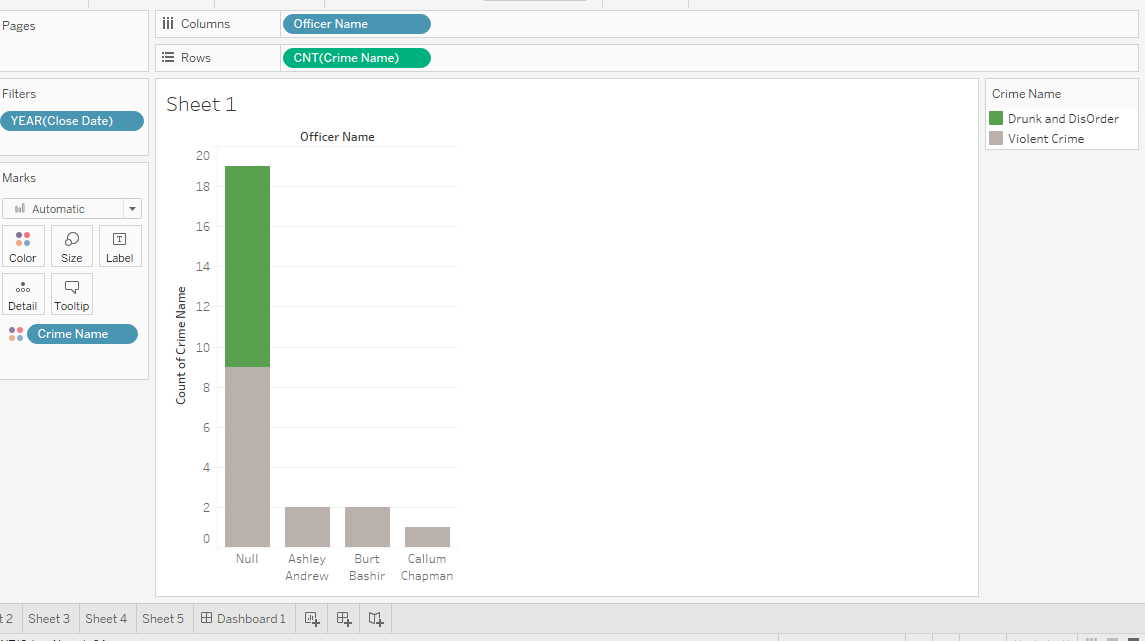


Chart 5 shows year and month along with total number of crimes where crime has been closed the most and the least

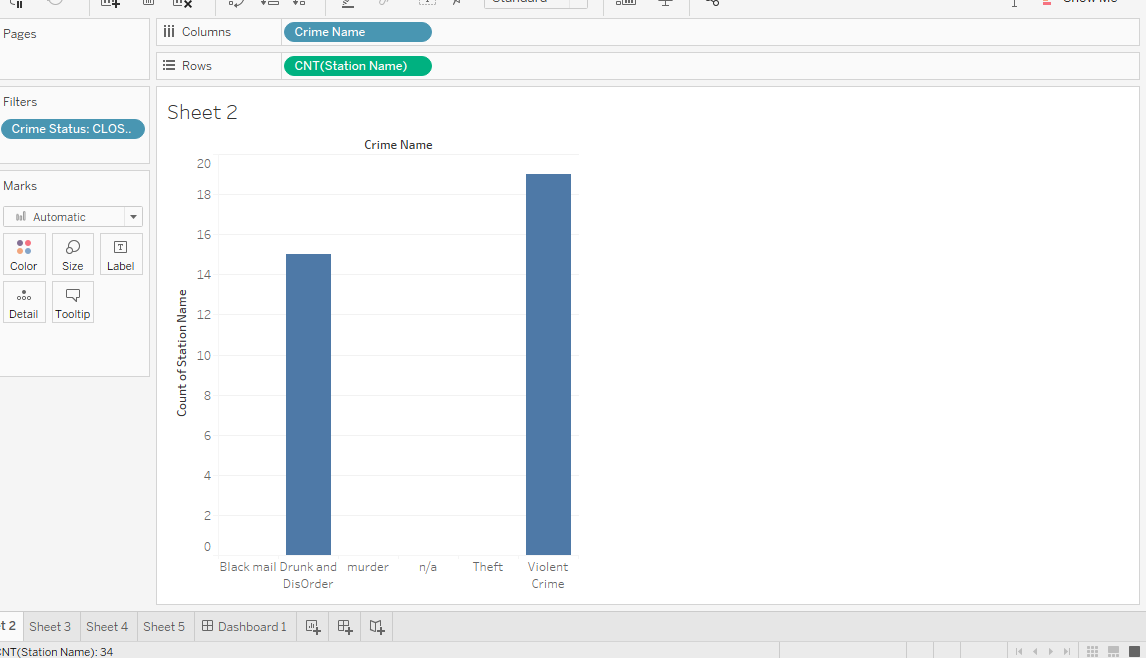
### Tableau Reports and Charts

The reports provided below were generated using **Tableau** software and are based on the data that were provided to us to carry out this task.

#### Report 1: Number of crime and crime name that has been closed by the officers per year

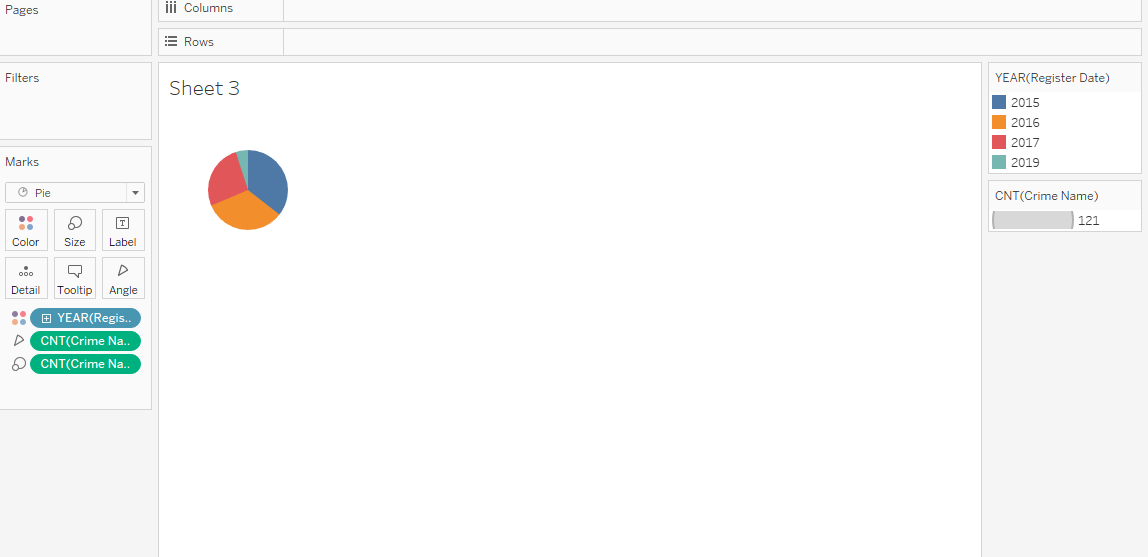
 The chart shows the name of officers along with crime type and number of crimes closed by them per year.

#### Report 2: Number of station and crime name where crime has closed



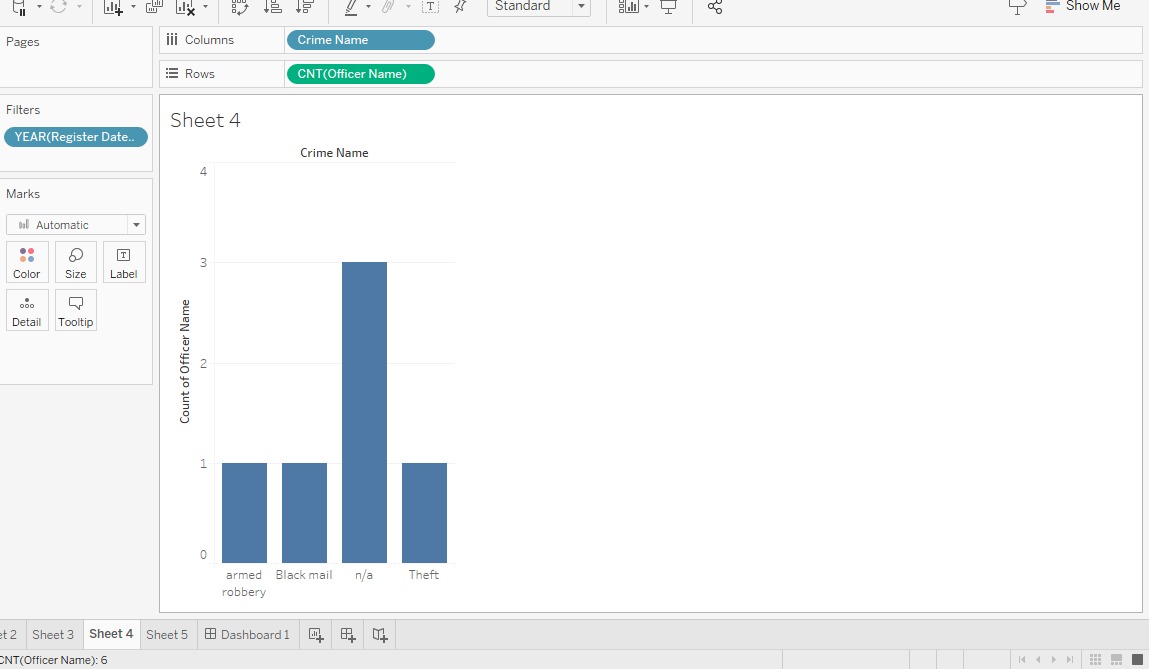
The chart shows number of station and crime type where crime has been closed.

#### Report 3: Number of crimes per year



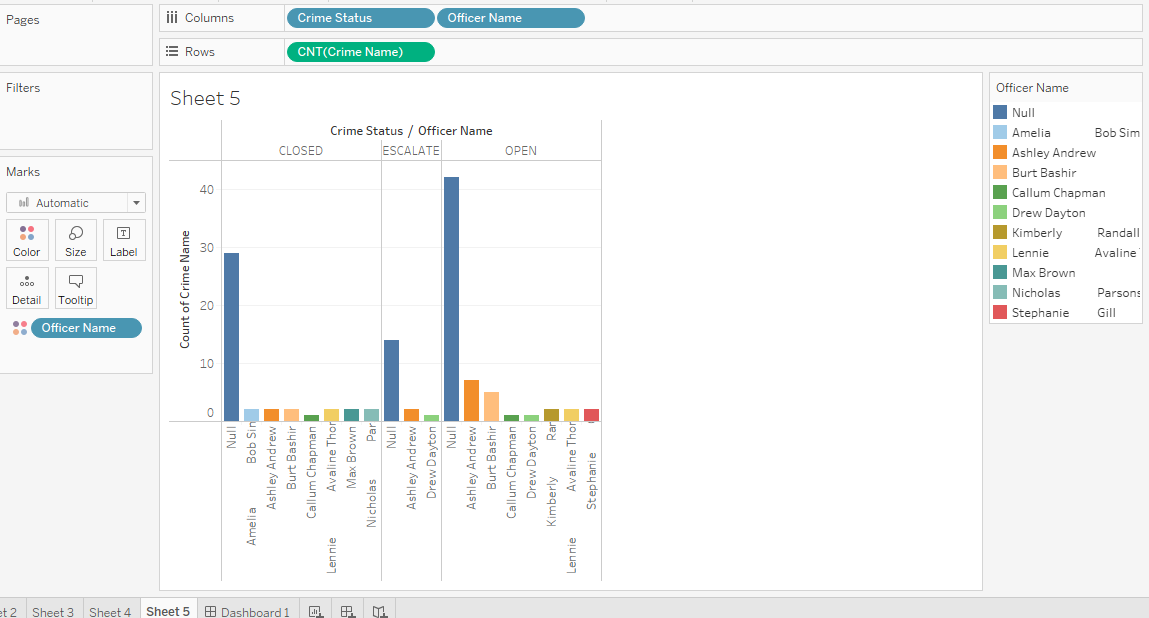
The chart shows total number of crimes per year

#### Report 4: Number of officer per crime type per year



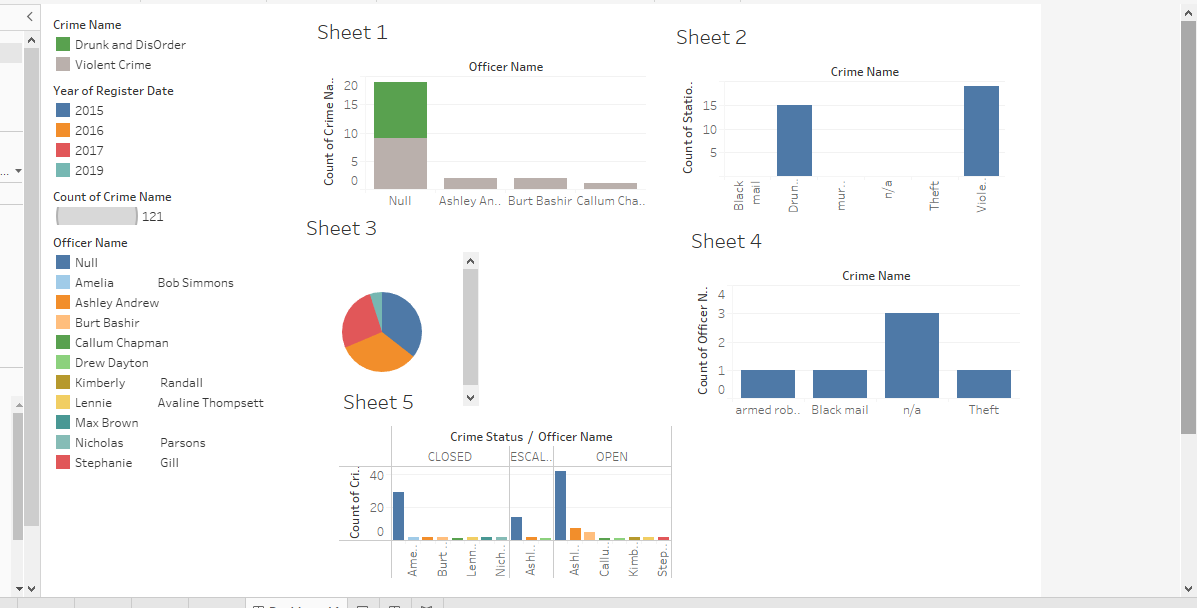
The chart shows officers along with crime that they have handled per year.

#### Report 5: Number of crimes per crime status per officer



The chart shows officer and number of crimes per crime status.

**Tableau Dashboard**

****

In this way, three different technologies were used for creating the reports and a proper data analysis and visualization was done using these tools for the case study.