Project Report

|  |  |
| --- | --- |
| **Course Name (NICF)** | Advanced Certificate in Data Science |
| Product Name (Marketing & Sales) | Advanced Certificate in Data Science |
| **Module Name (NICF)** | WSQ- Data Analysis and Visualization(SF) |
| Product Name (Marketing & Sales) | WSQ- Data Analysis and Visualization(SF) |

|  |  |  |  |
| --- | --- | --- | --- |
| Student name | | Assessor name | |
| Tan Tiong Choon (Chen ZhongChun) | |  | |
| Date issued | Completion date | | Submitted on |
| 27 Sep 2023 | 11 Oct 2023 | | 11 Oct 2023 |
|  | |  | |
| Project title | Russia\_Ukraine: The War Analysis and Visualization | | |

|  |
| --- |
| Learner declaration |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.    Student signature: Date: 11 Oct 2023 |

**Project Overview**: Describe the Project along with Project Outcomes (Explain the Project in your own words in 15 – 20 lines)

This project requires us to collect, verify, analyse and showcasing the losses of equipment and personnel involved in the on-going war between Russia and Ukraine. It will provide us with the data-driven understanding of the cost and losses involve in the conflicts and the impact on both nations. It also enables us to have an overall view of the real-world situation of a data scientist works.

We will need to collect, verify and analyse data related to equipment and personnel losses in Russia and Ukraine conflicts, thereafter create visually informative data visualisation using charts to present the losses effectively. These will enable us to identify the trends, patterns, correlations and disparities in equipment and personnel losses between the two nations over time.

The expected outcomes will be showcasing a comprehensive dataset detailing equipment and personnel losses in Russia and Ukraine during the conflicts, with a focus in accuracy and reliability. We shall be presenting a set of informative data visualisation that will assist the stakeholders, policymakers, researchers and the public understand the scale and impact of the conflicts. With this in-depth analysis of the trends, patterns, correlations and disparities, we shall be able to present the audience with better understanding of the dynamics involved, identify the potential area for international humanitarian assistance, increase awareness and empathy for the human suffering and material costs suffered by both nations, support diplomatic efforts to find peaceful resolutions to the on-going conflicts and reduce further losses endured by the both nations.

Project Technical Environment: (Describe the Architecture with Tools used)

We shall be using Python programming with Jupyter Notebook software for scripting.

Python is the primary programming language used for data analysis and visualization. We will use Python for data manipulation, analysis, and visualization.

Jupyter Notebook allows us to combine code, visualizations, and documentation in an interactive environment.

The data will be provided by the course which contains 2 datasets file containing the equipment and personnel losses for Ukraine and Russia.

We will use Pandas and NumPy libraries to collect, extracting, verify and analyse the data and ensure that the data is in a suitable format for analysis.

For data visualization, we will be using Seaborn and Matplotlib. We will be presenting line plots, bar charts, histogram, scatter plots and heatmaps, to visualize losses over time or by category.

1. Analytical Technique & Tools used: Describe the Analytical Technique and Tools used in the Project

The integrated development environment we will be using is Jupyter Notebook for coding.

We will be using Pandas, NumPy, Matplotlib and Seaborn library to collecting, extracting, cleaning, analysis and creating data visualization.

We will be using Pandas and Numpy to read the file, display or describe the file contents or format or shape, calculate the mean, sum, count, standard deviation, drop the null values, create calculative columns and sort the data for losses in equipment and personnel for both Ukraine and Russia.

We will use Matplotlib and Seaborn to create line plots, bar charts, histogram, scatter plots and heatmaps to identify trends, compare losses, seasonality, and irregularities in losses over time or by category and explore correlations to determine if there is a relationship between these variables.

1. Activity 1: Task Summary

We will be using Python programming and Jupyter scripting language to collecting, extracting, cleaning, analysis and creating data visualization.

The flow is as follows:

**Requirement Collection**

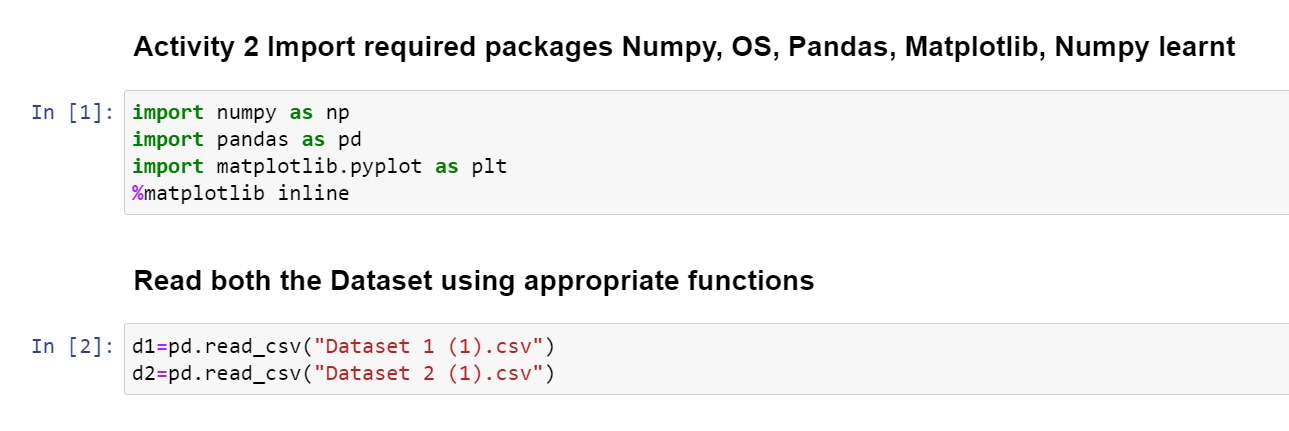
**Collection and Extracting Data**

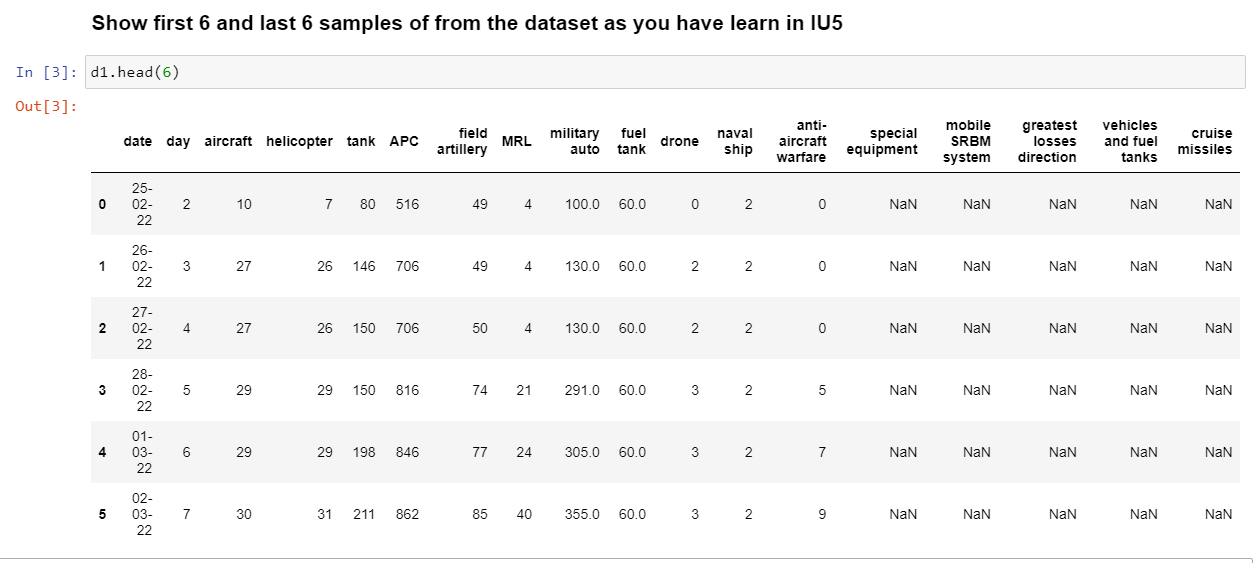
**Data Cleaning**

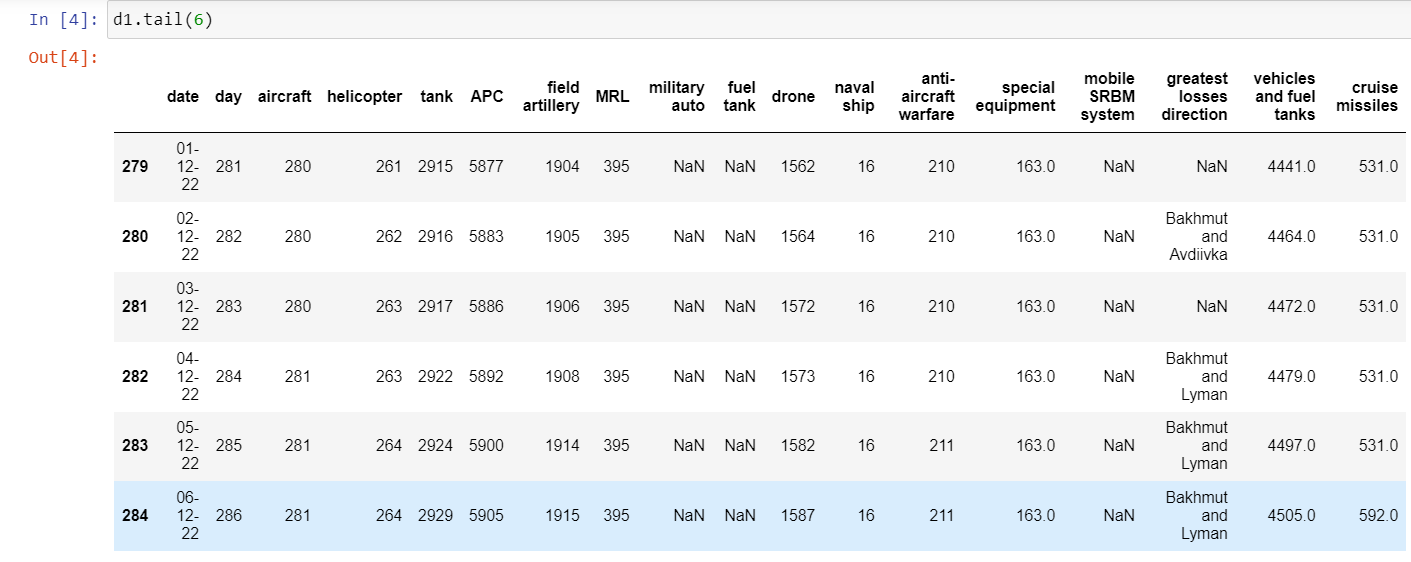
**Data Analysis**

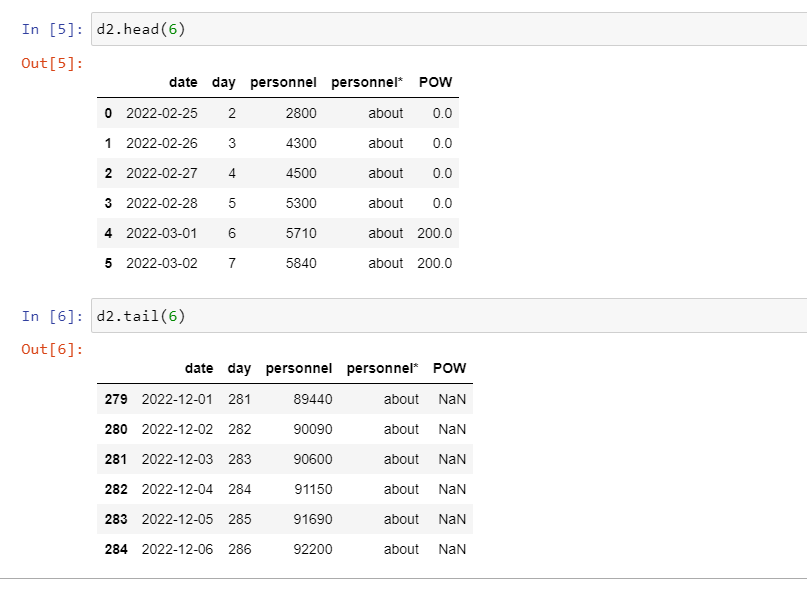
**Data Visualisation**

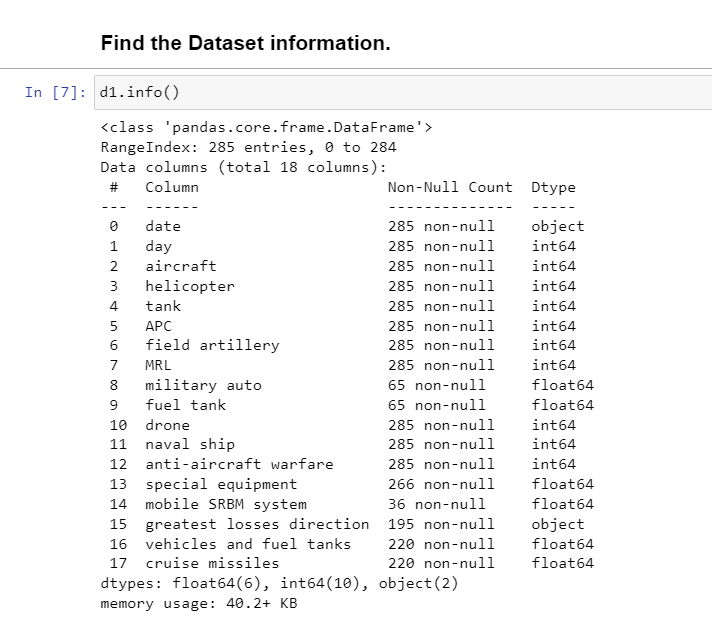
1. Activity 2 and Screenshots :

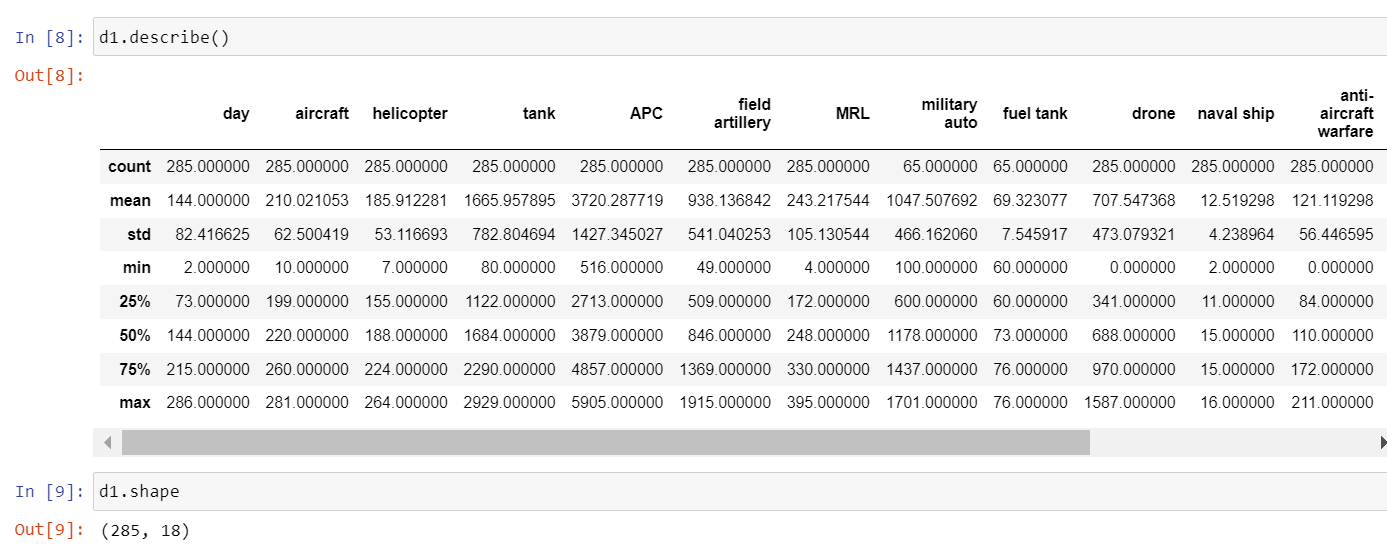


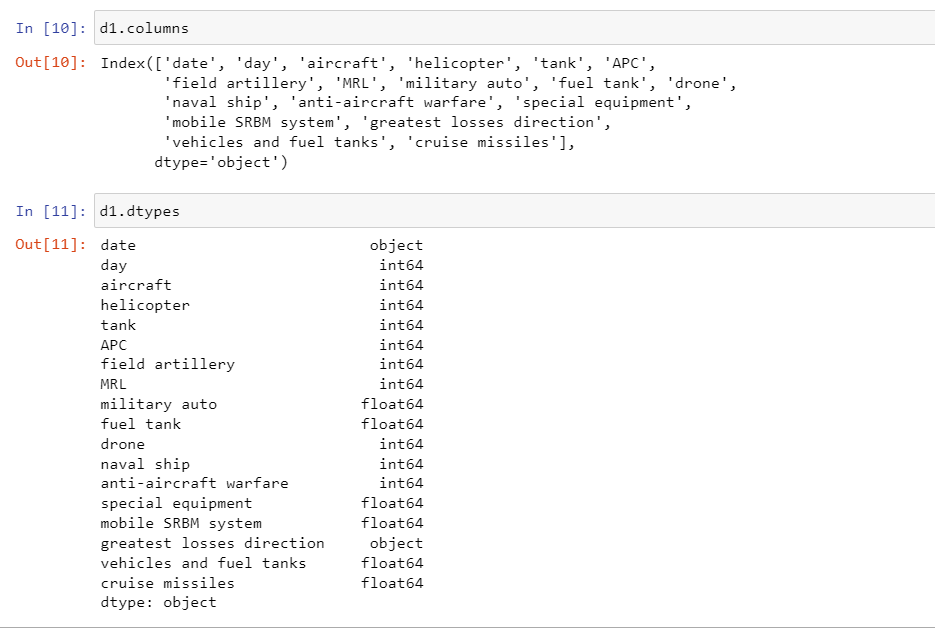


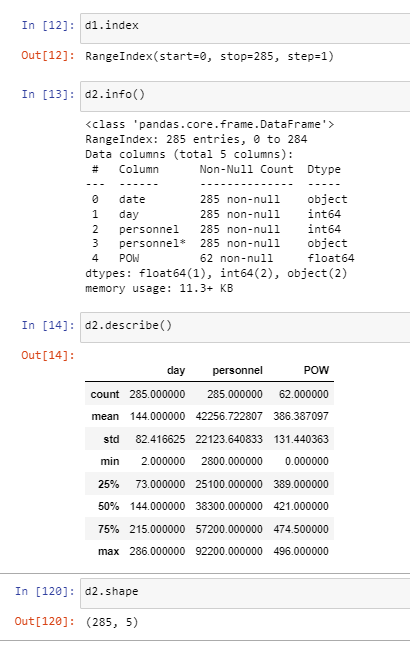


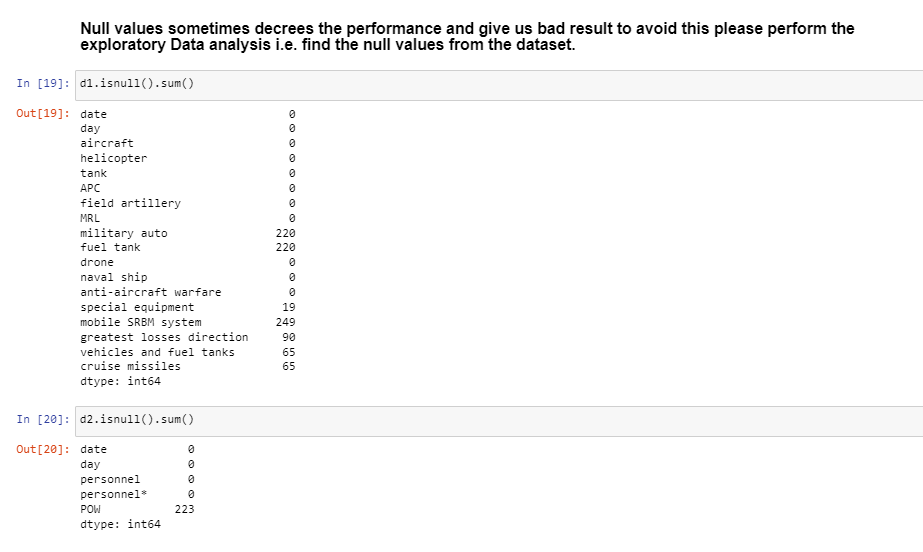


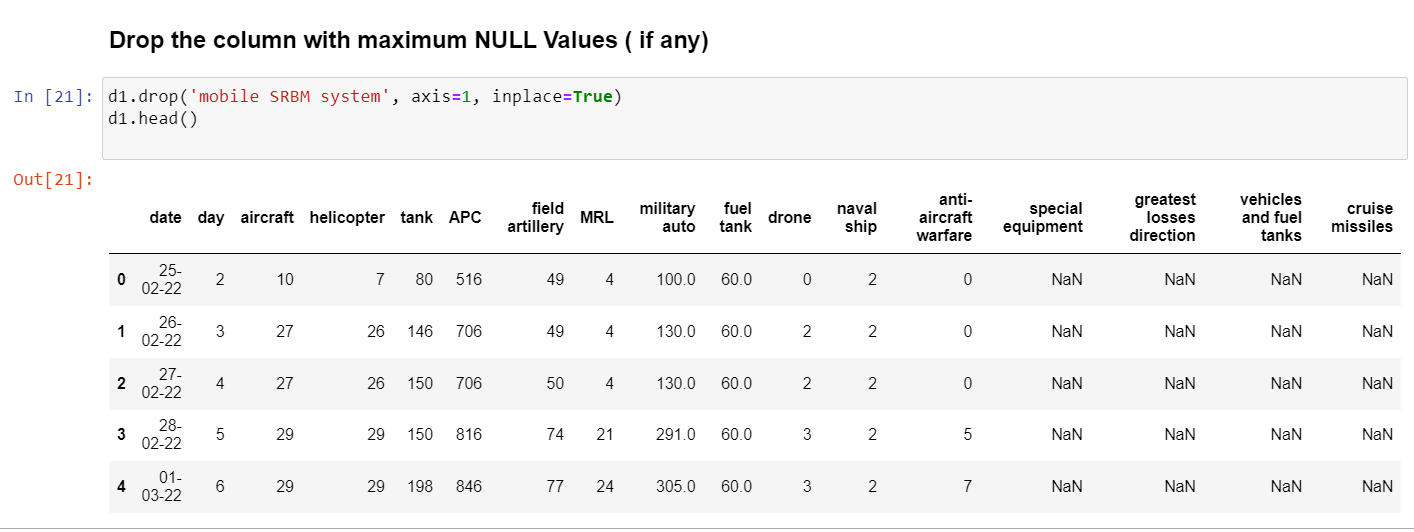


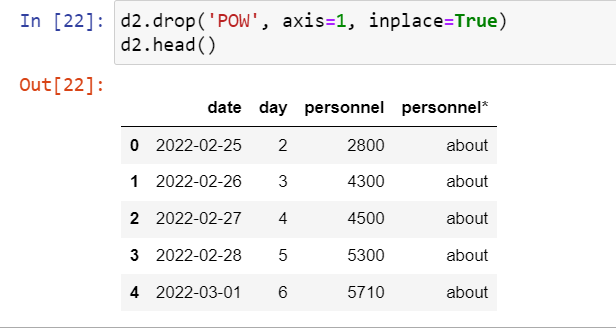


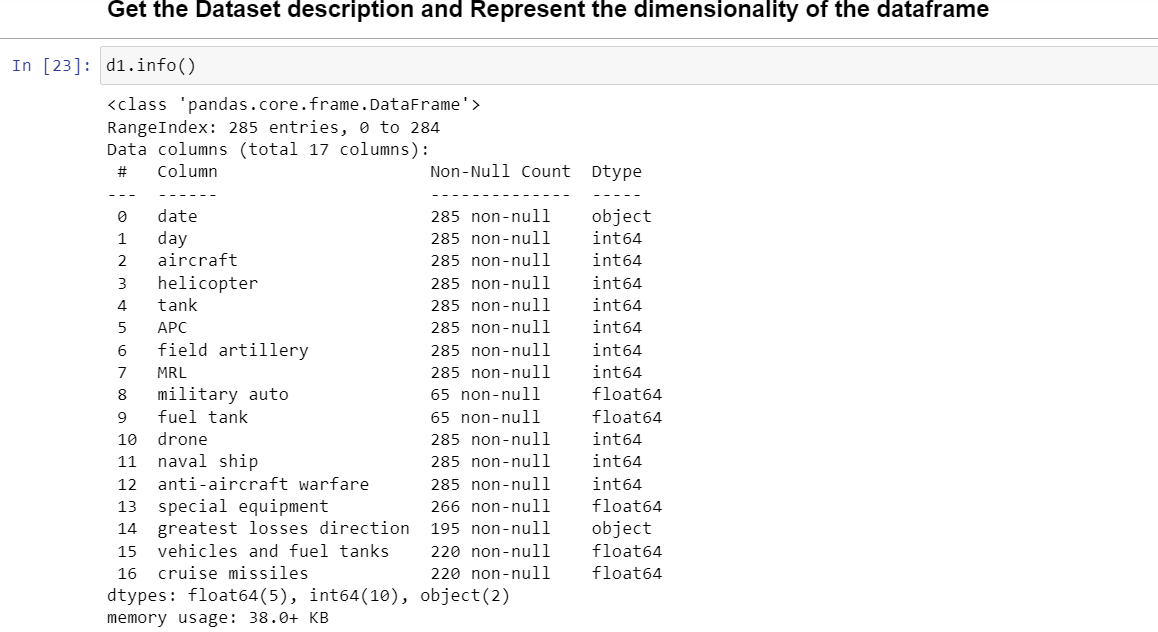


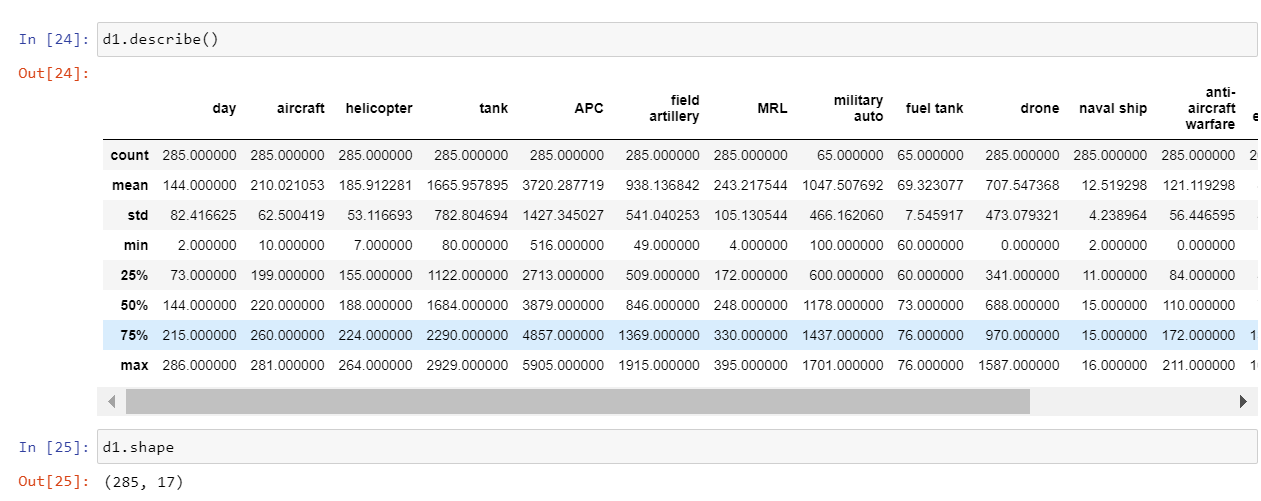


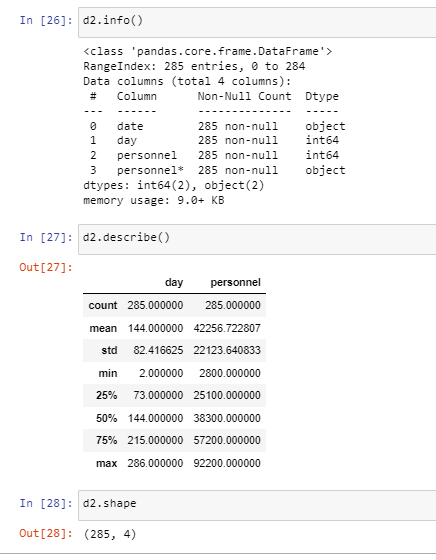




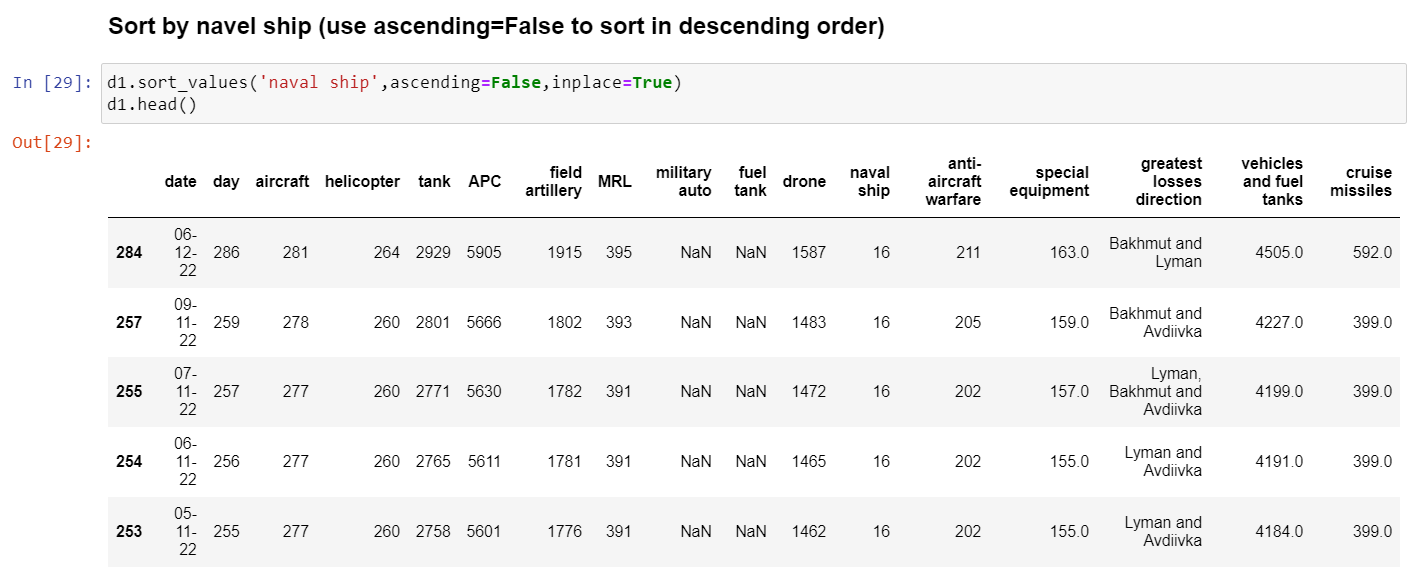




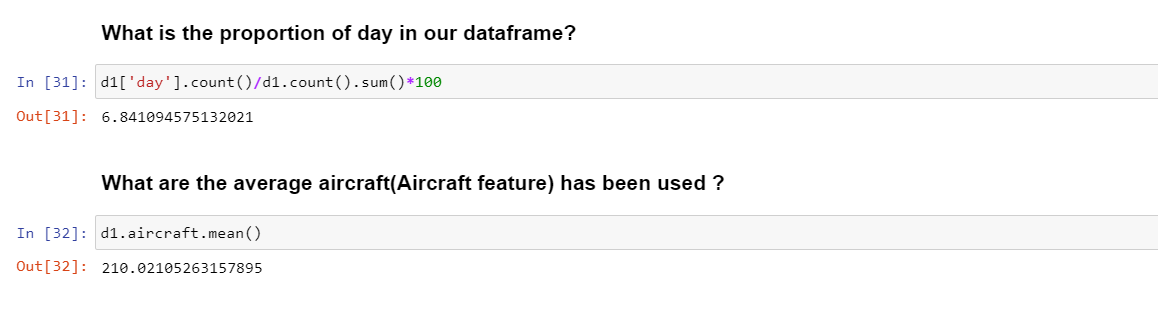


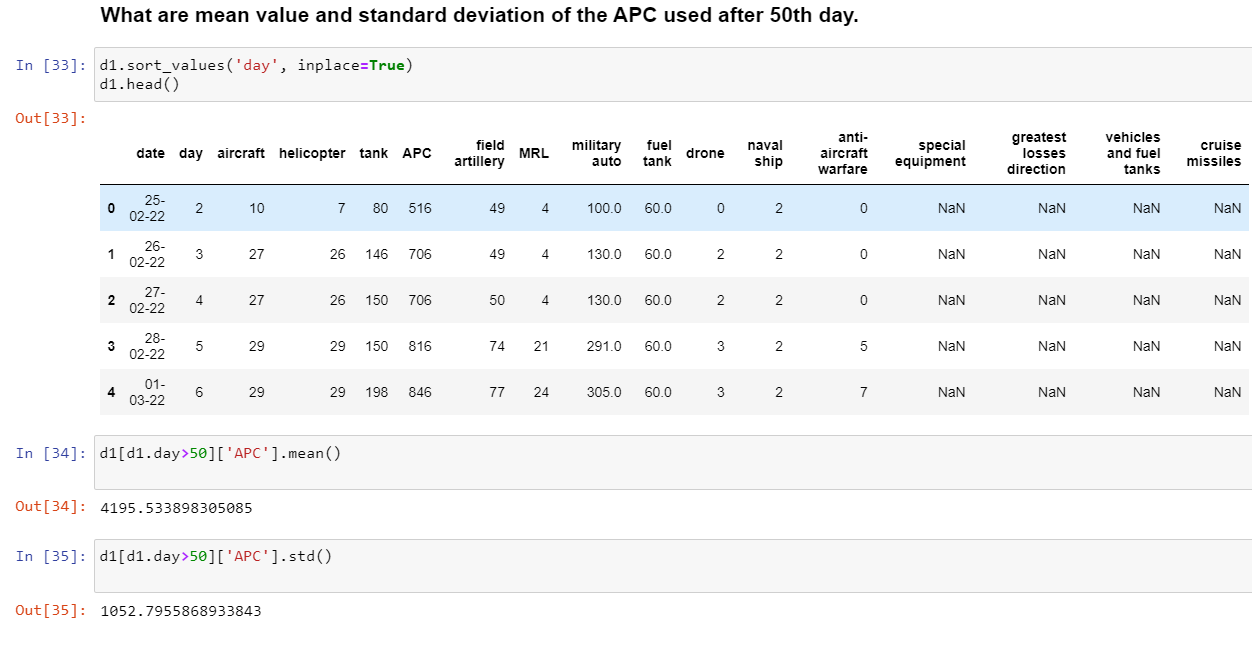


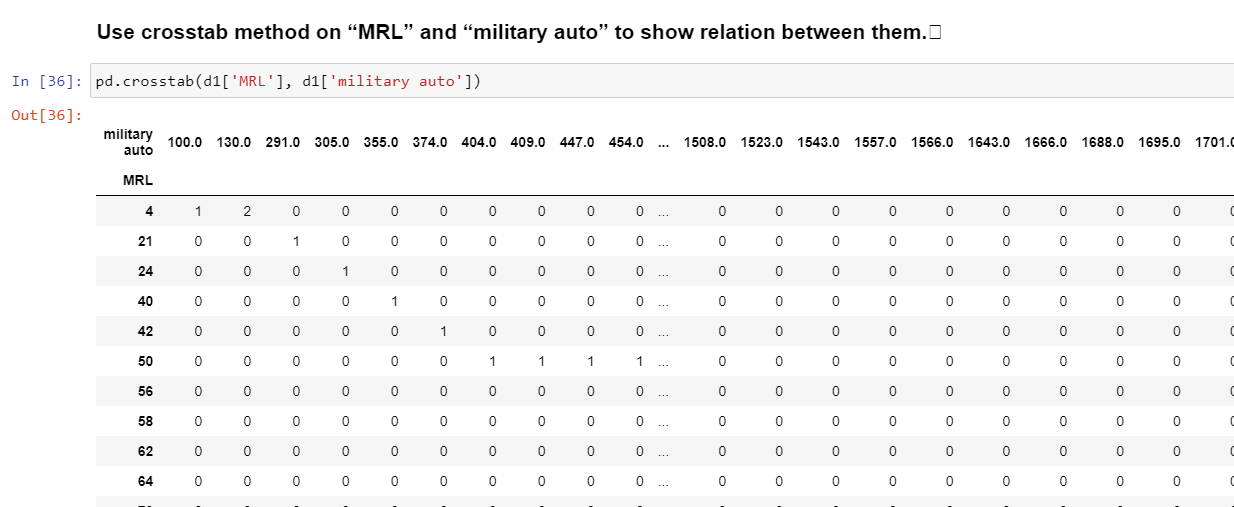
1. Screen-shots of each task of Activity 3

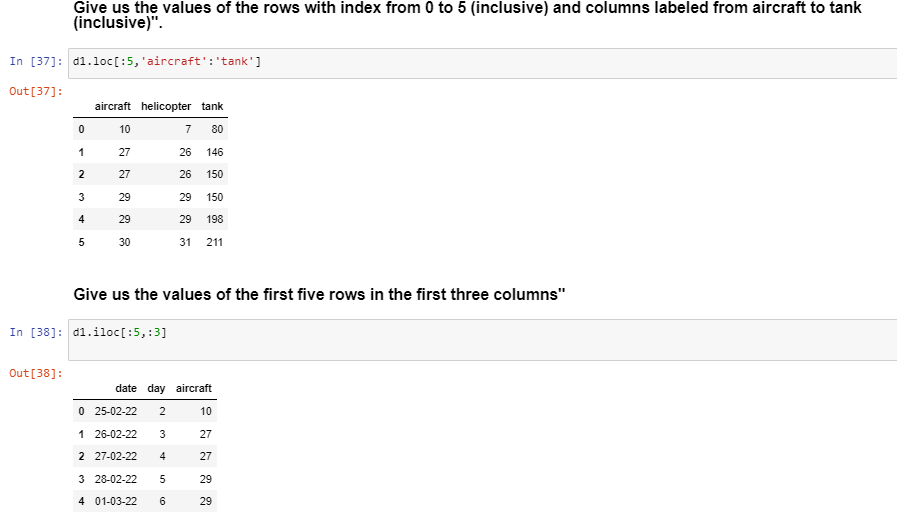




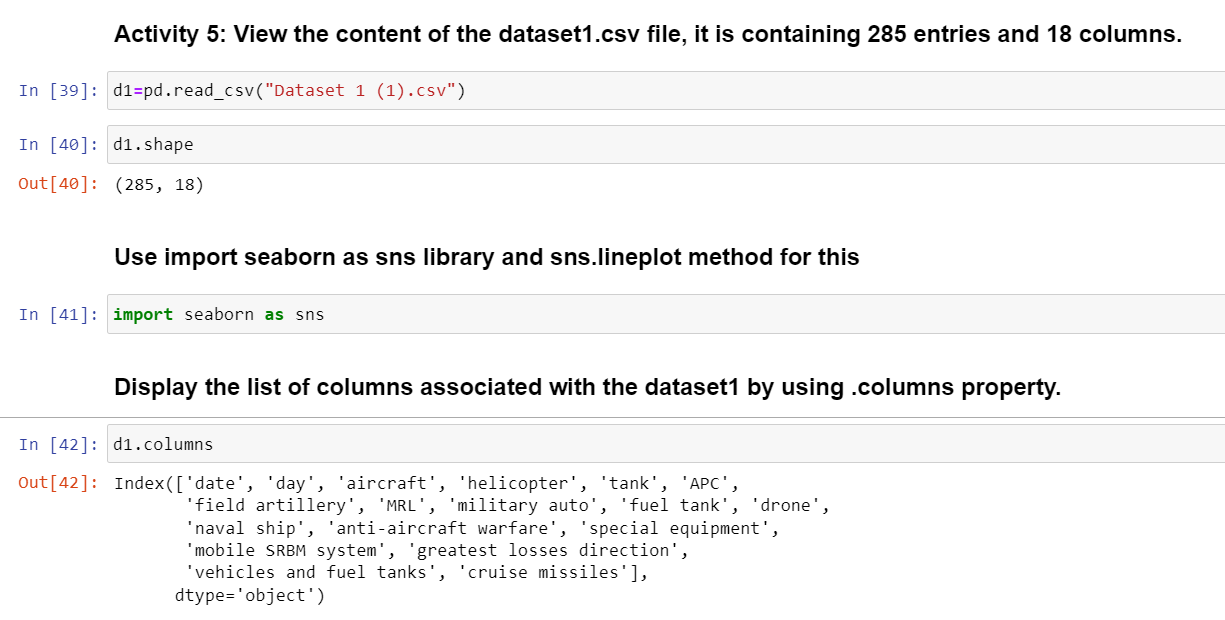


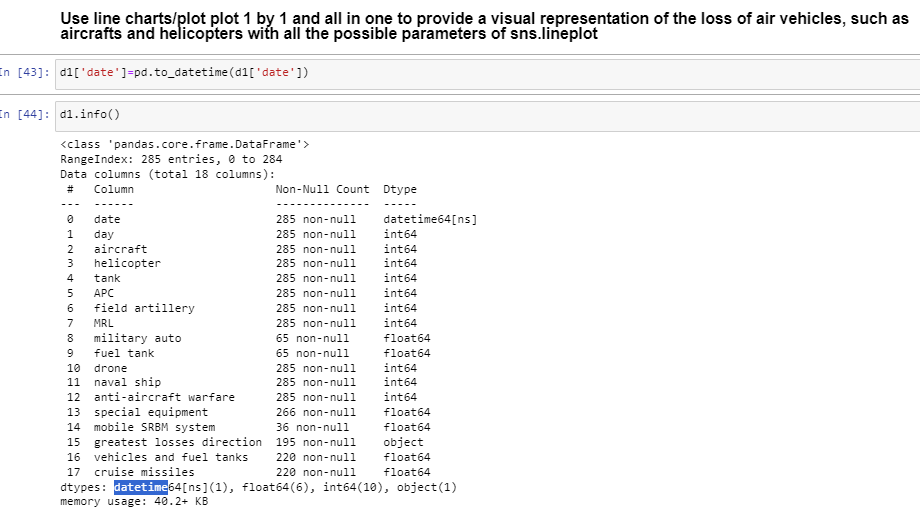


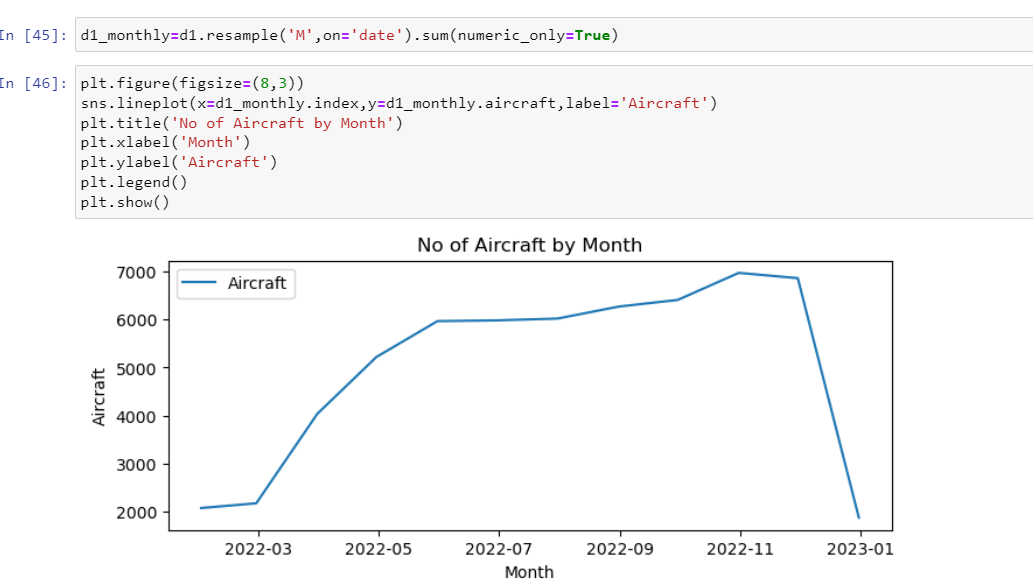


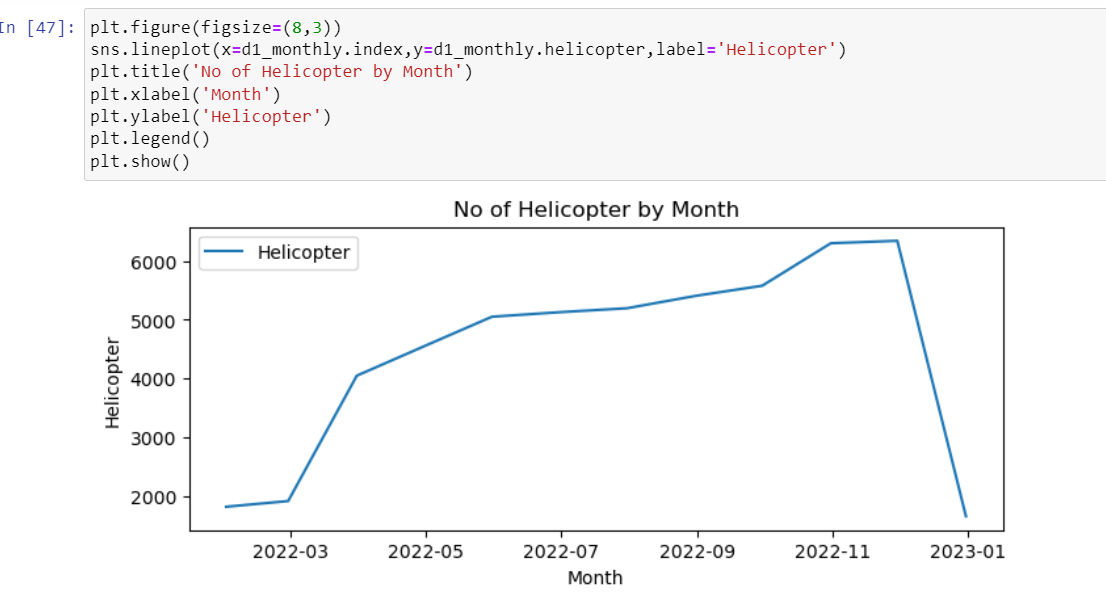


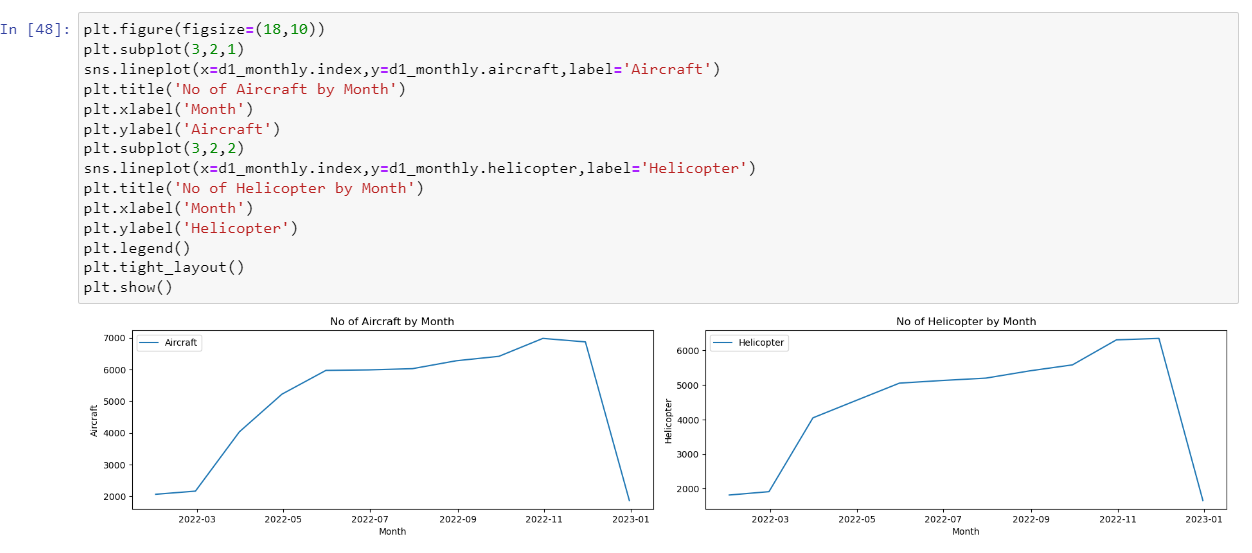
6. Screen-shots of each task of Activity 5

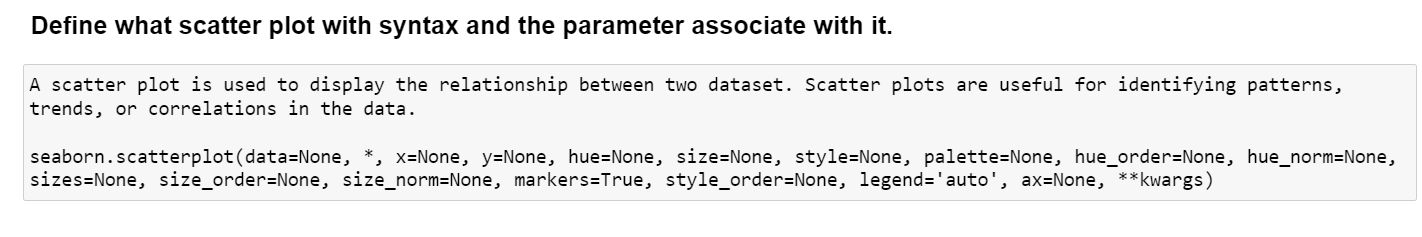


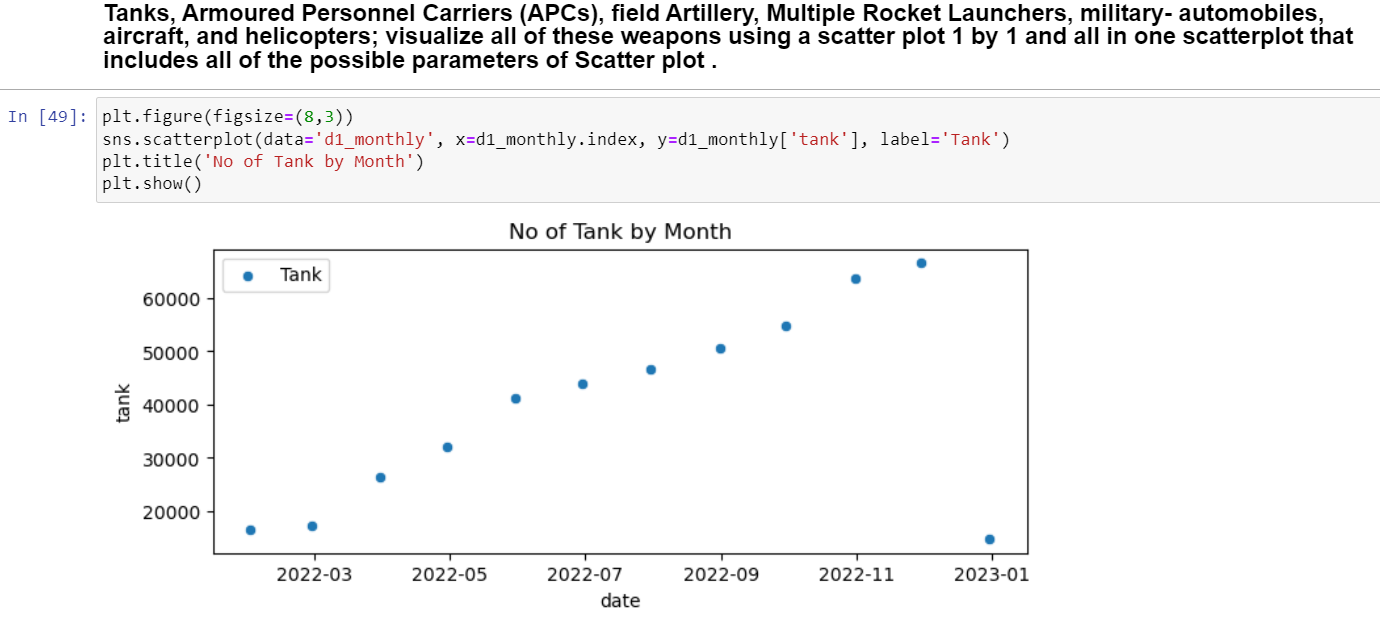


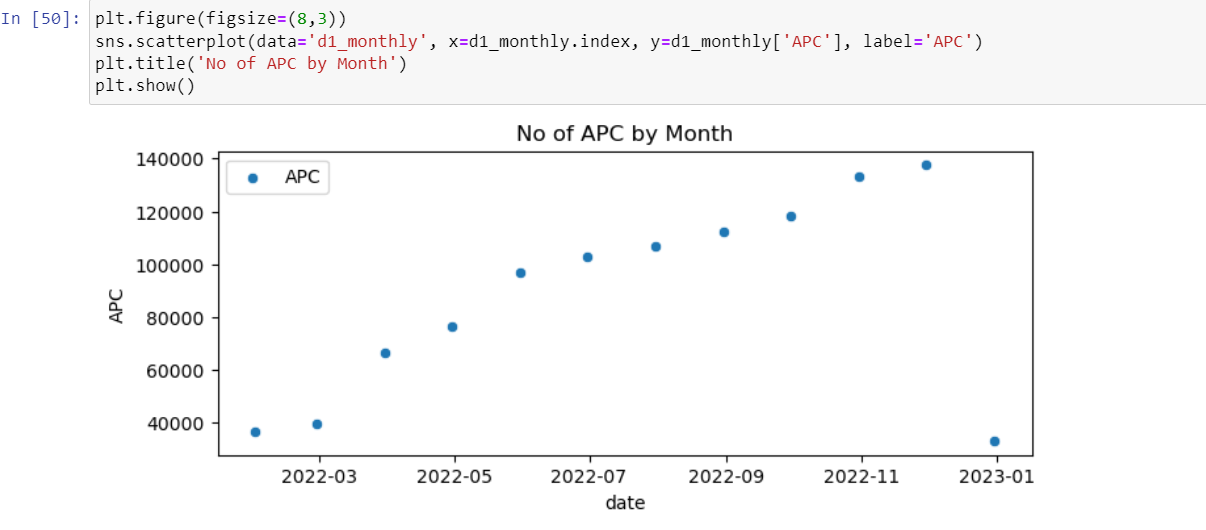


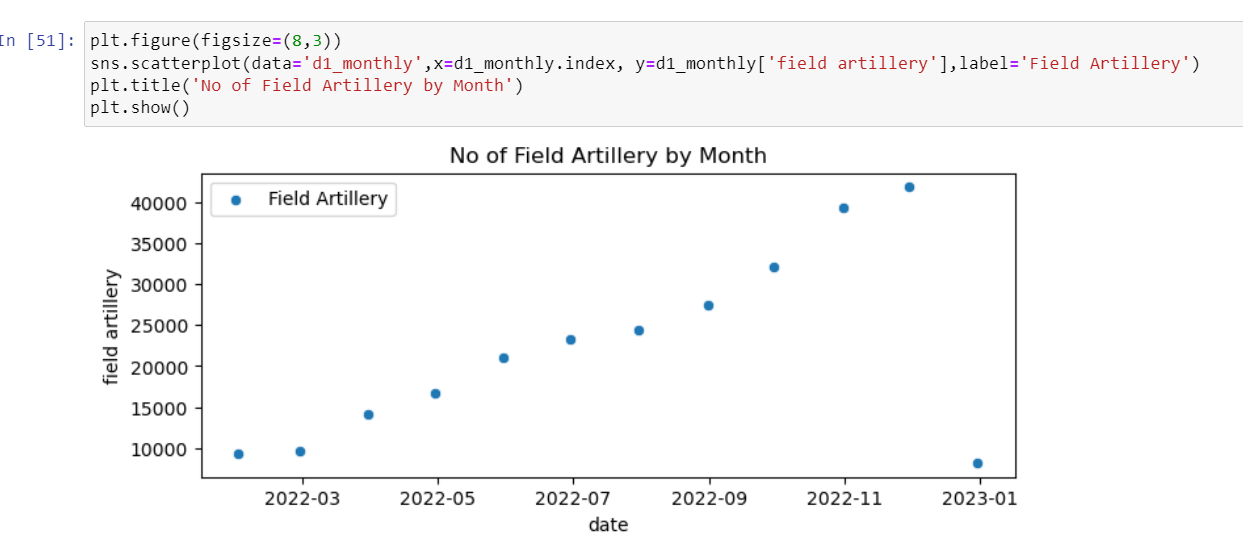


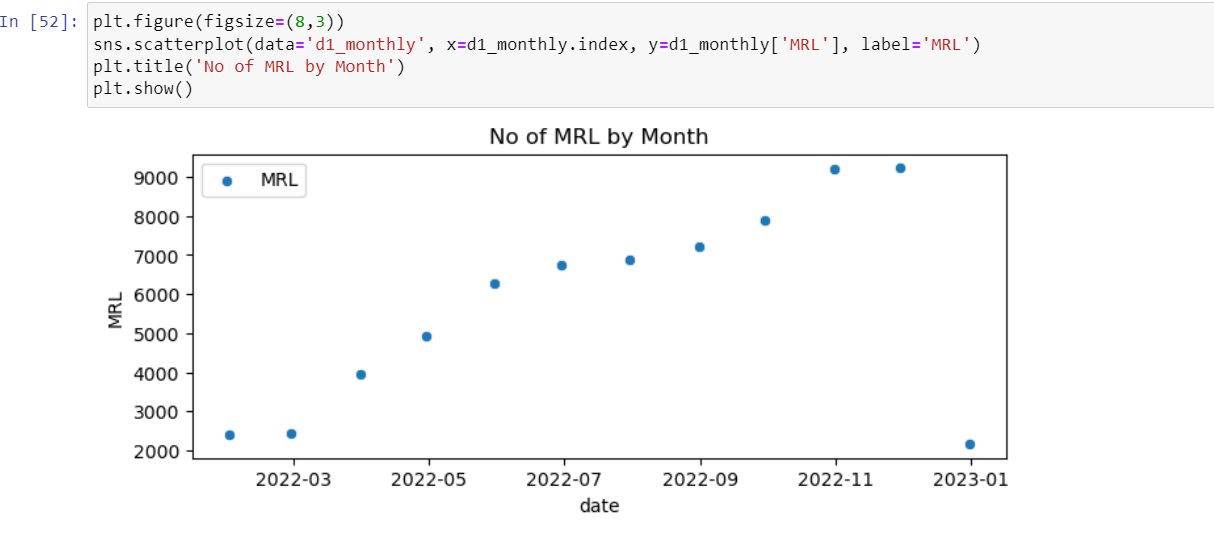


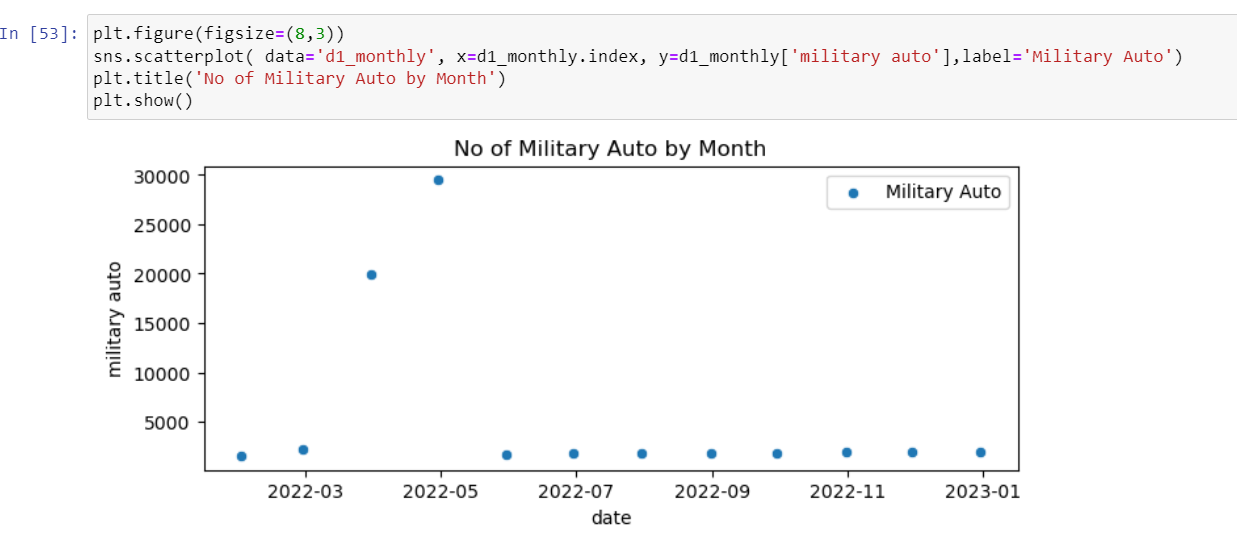


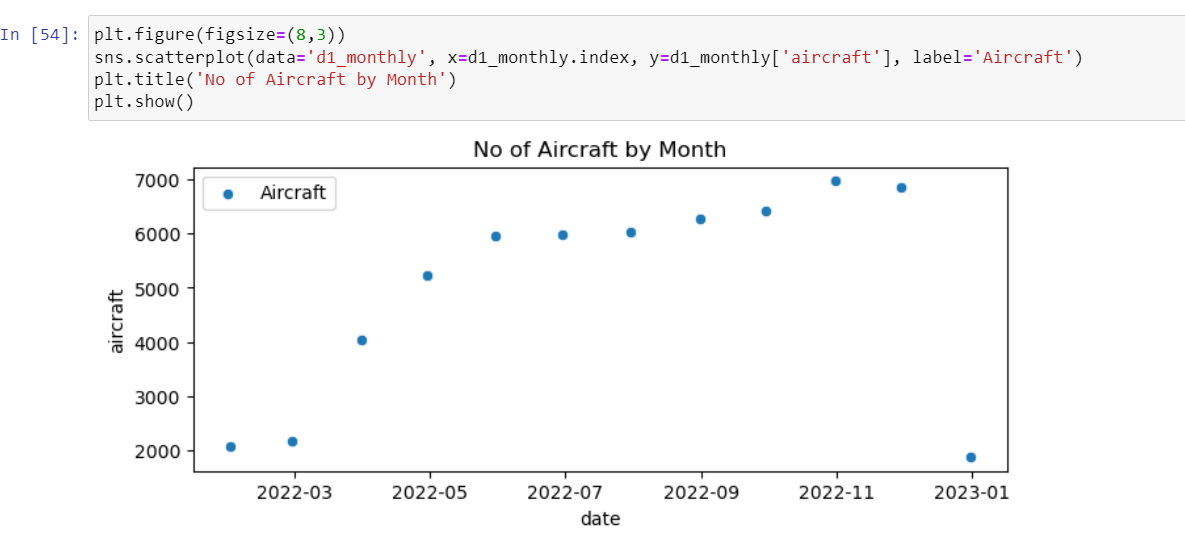


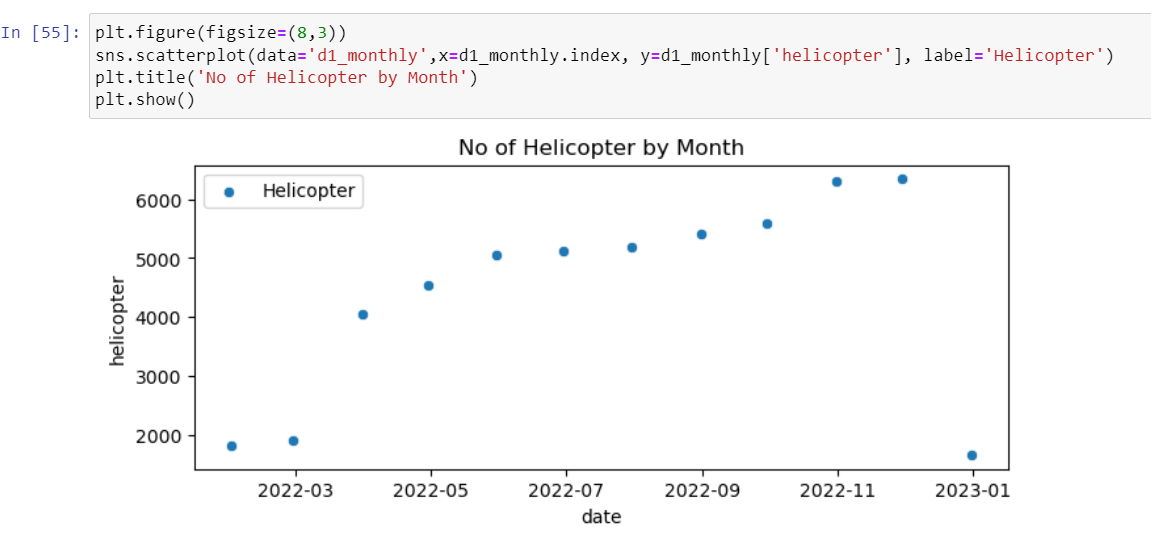


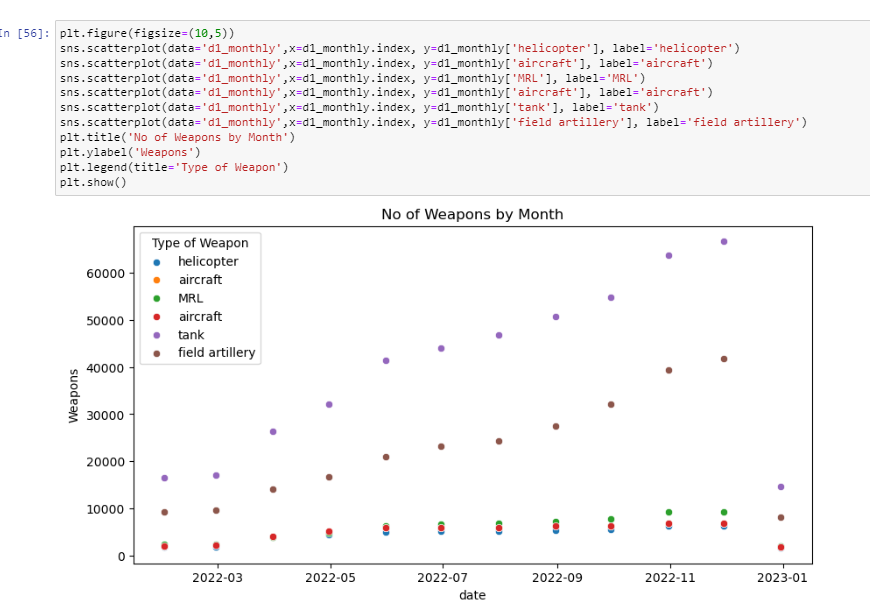


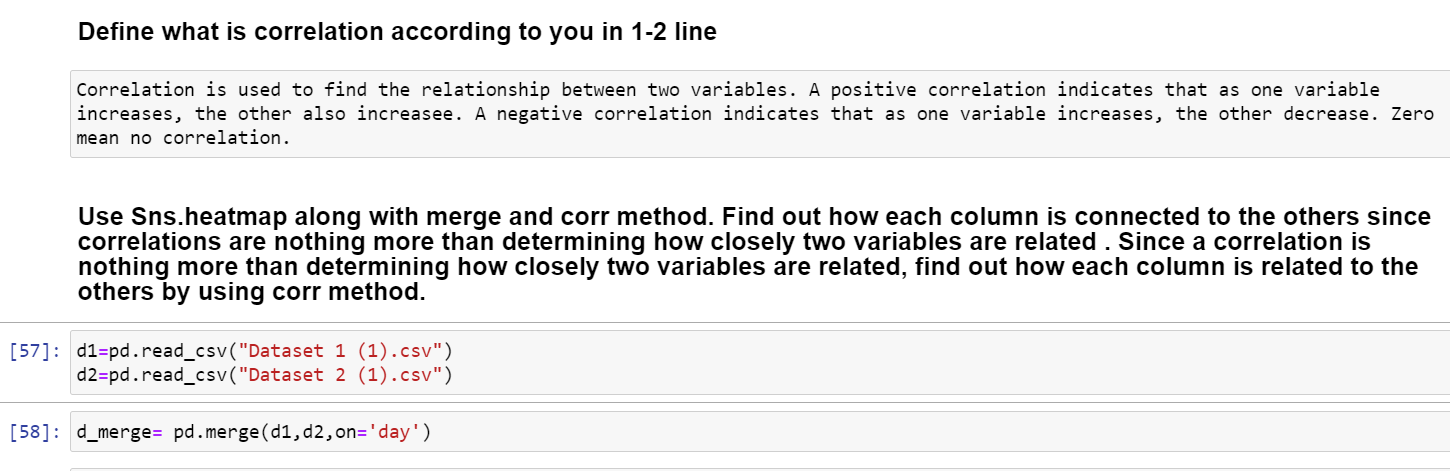


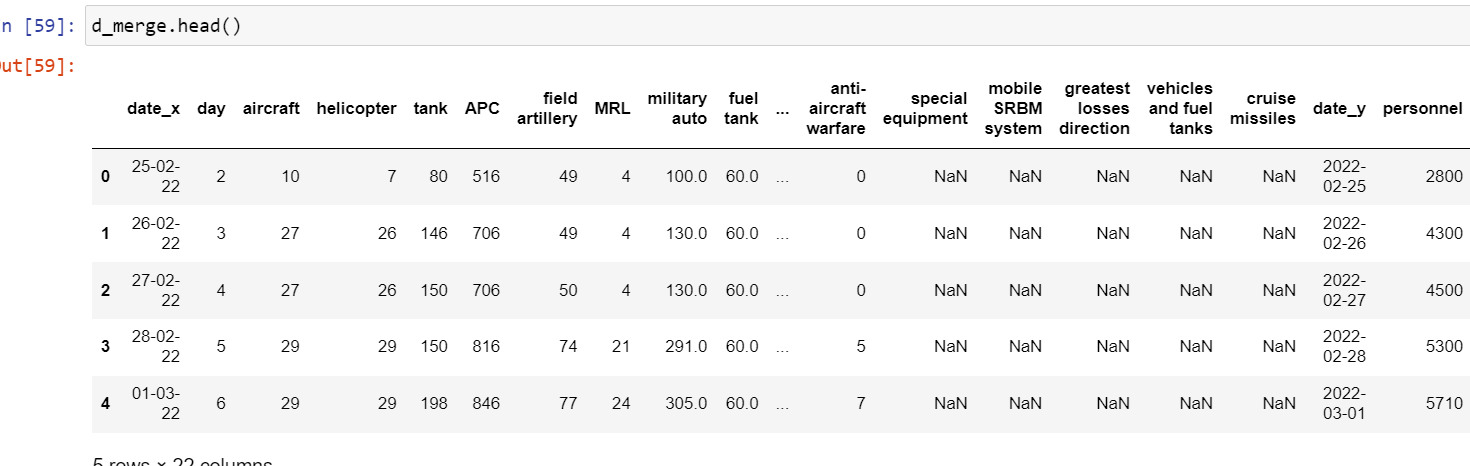


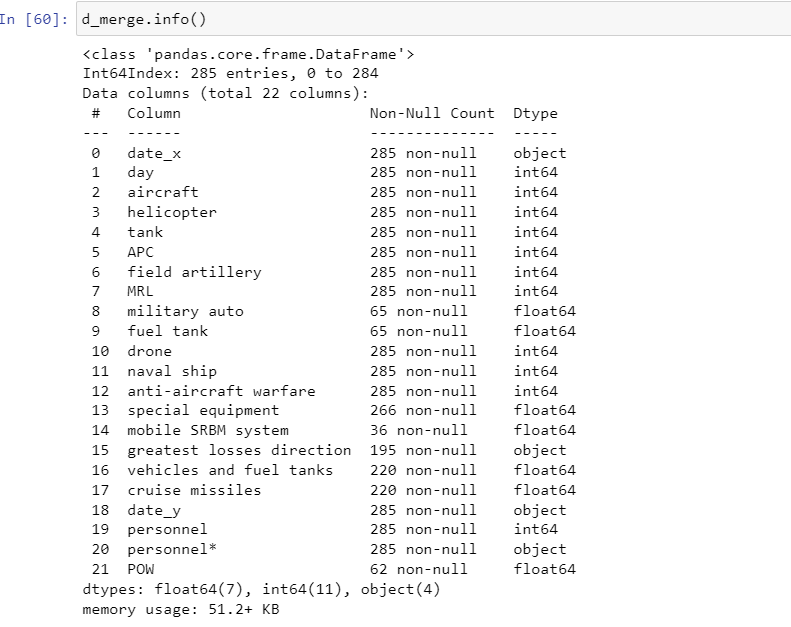


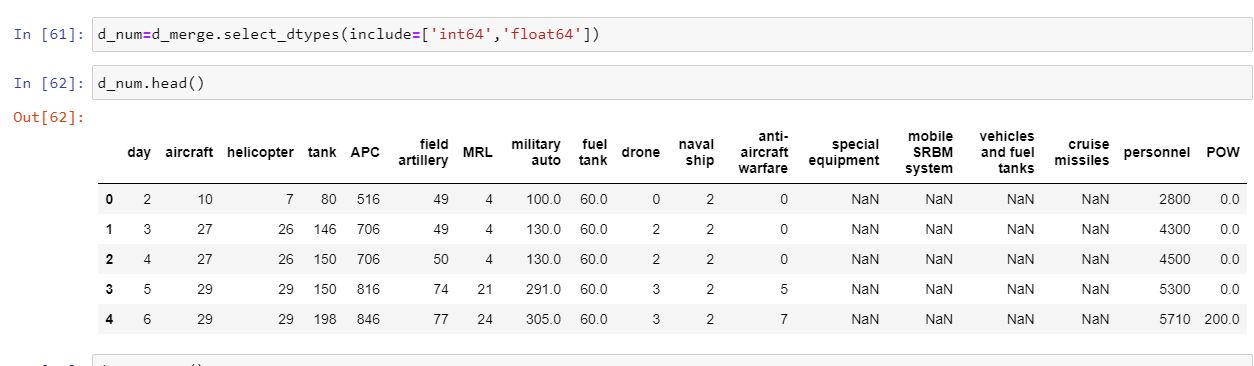


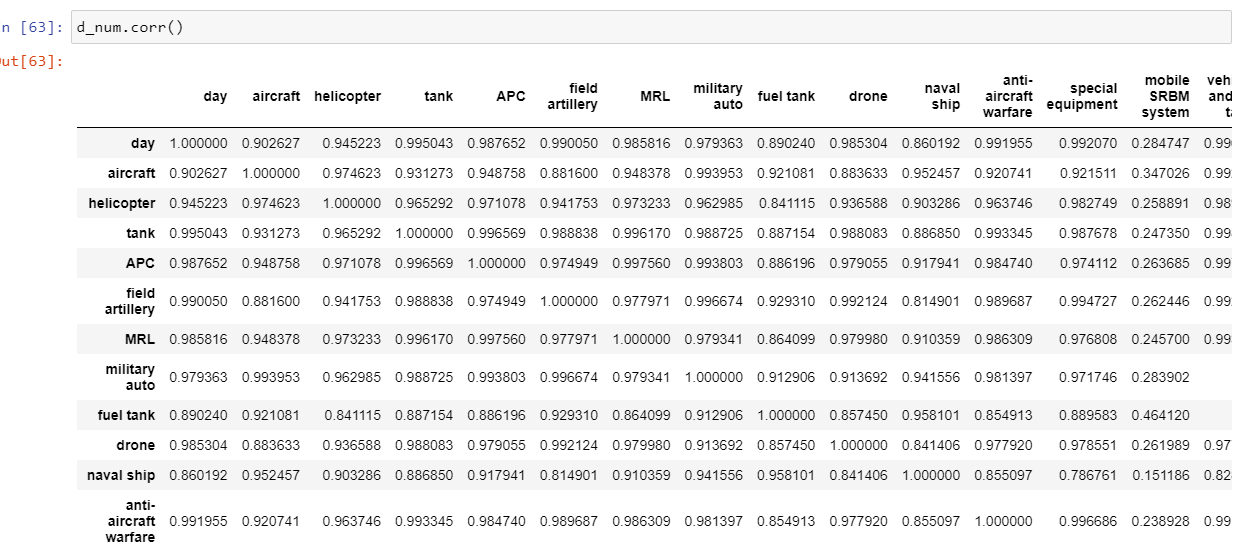


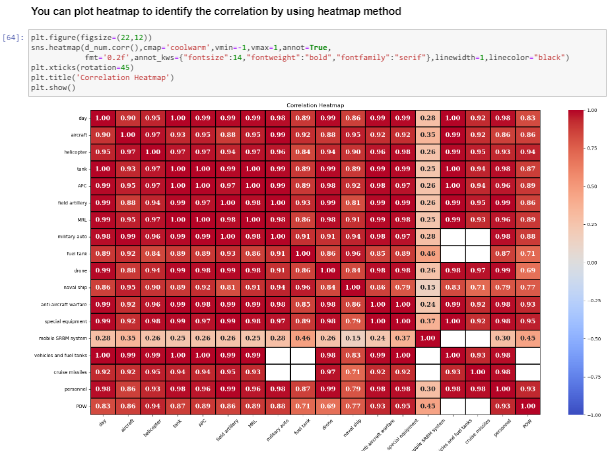


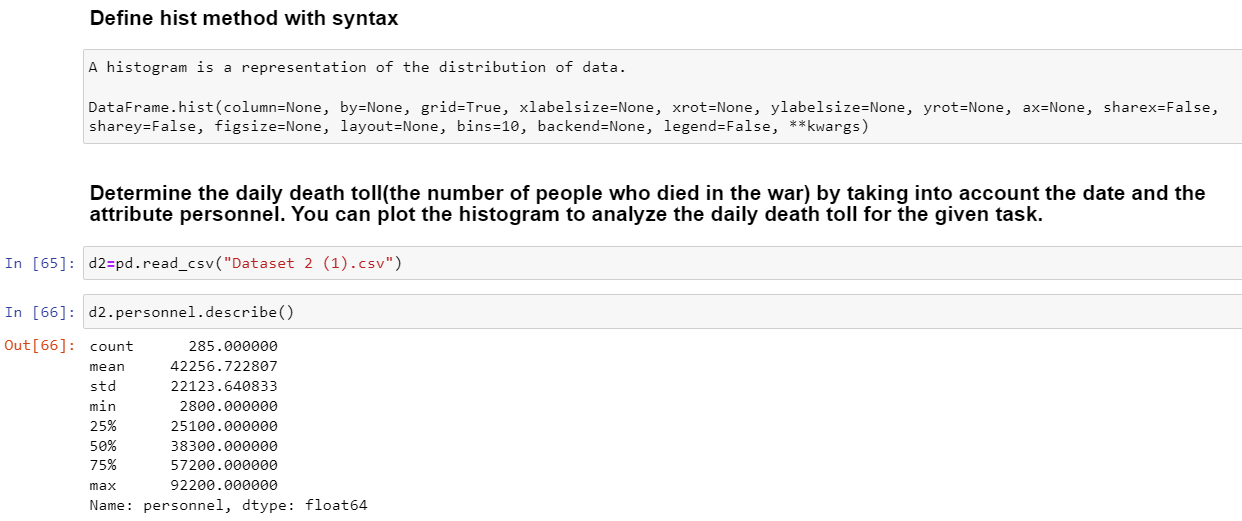


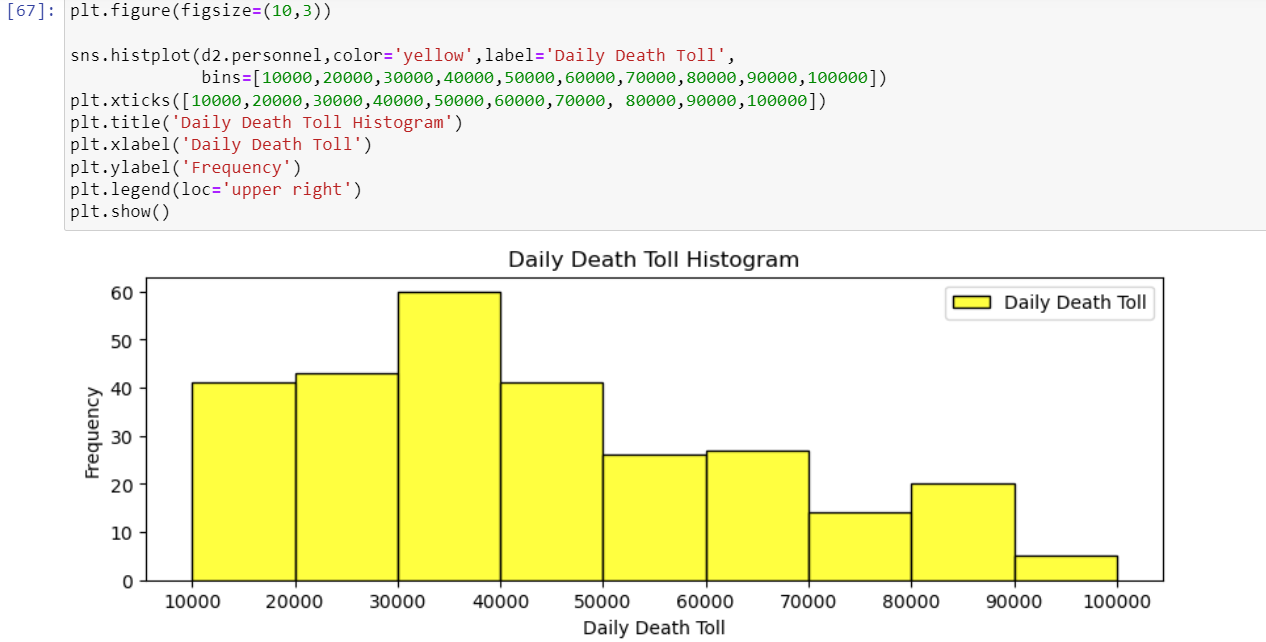


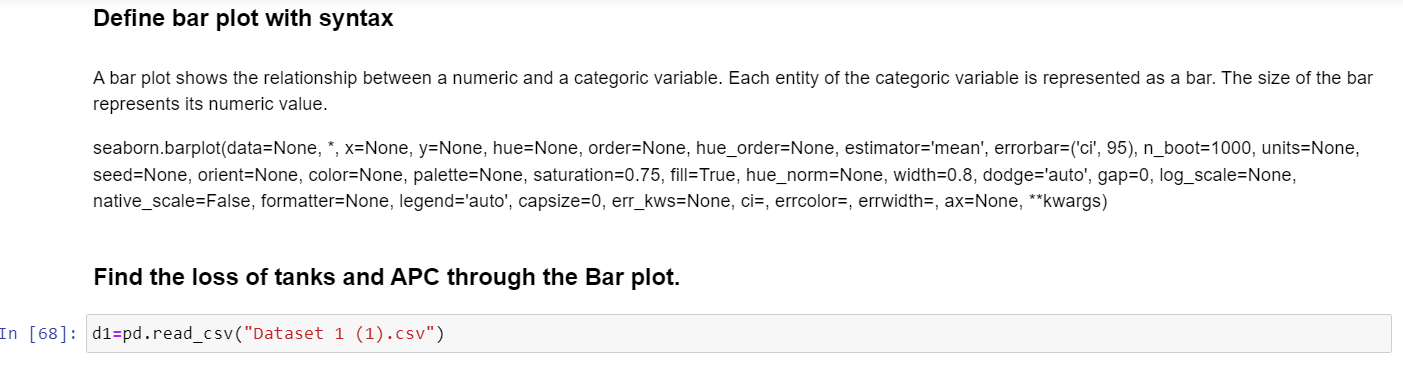


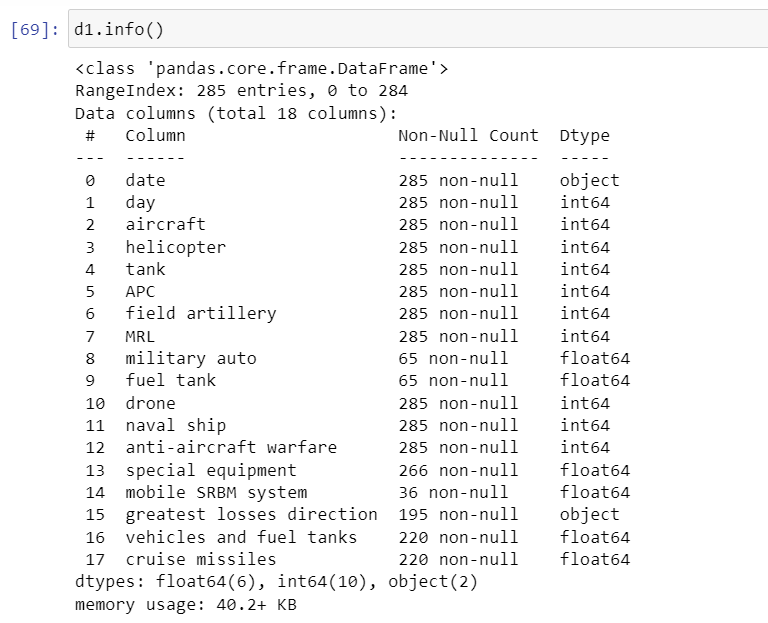


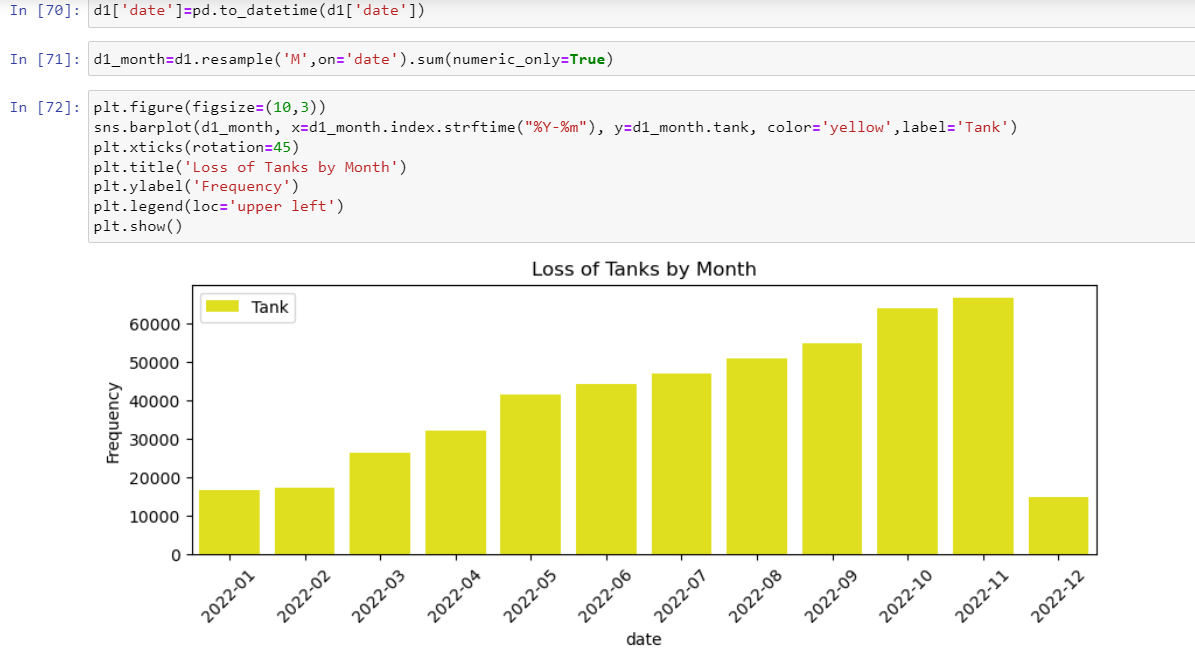


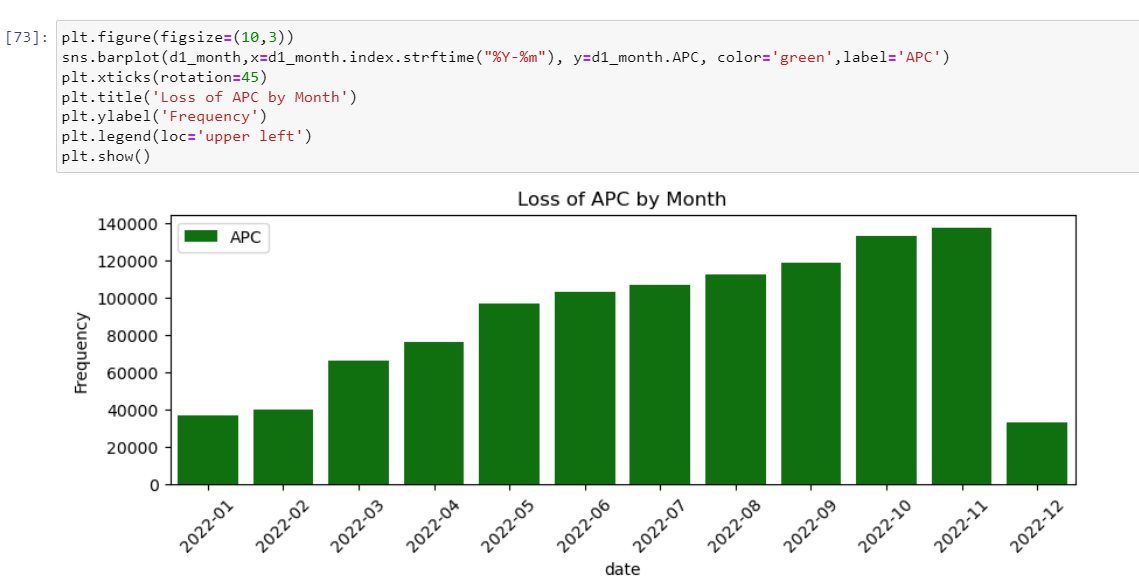












7. Annexure 1

* 1. Script1

1. Annexure 2
   1. Script 2
2. Annexure 3
   1. Script 3
3. Annexure 4
   1. Script 4