Task 4.1P

Visualize data using matplotlib

Pie chart and bar chart creation with `matplotlib` and data processing with `pandas` were used in the cybersecurity attack data visualization work. Functions like groupby(), sum(), and value_counts() must be understood in order to load the data and group it to count attack kinds. A clear comparison of the number of attack types across categories was shown by the bar chart, while the distribution of assault categories was effectively displayed by the pie chart, with the largest portion emphasized. By inferring labels directly from the data, the initial labeling and value issues were rectified. The practical uses of these abilities in cybersecurity and other data-driven professions were highlighted, along with the significance of proper data representation, in this exercise. Developing my knowledge of data processing techniques and investigating sophisticated visualization tools are among my future objectives.

```
In [28]: # Here I have read the  .csv file name
         file_name = 'attack-type-frequency.csv'
         data = pd.read_csv(file_name, index_col=0, engine='python')

         #Here I have used head to display the first five rows of the data frame.
         data.head()

         # Group by attack category and count the number of attack types and sum of attacks
         attack_types_count = data.groupby('category').size()
         attack_numbers_sum = data.groupby('category')['number_of_attack'].sum()

         # Define the colors and labels
         colors_bar = ['blue', 'red', 'green', 'yellow']  # Colors for the bar chart
         colors_pie = ['blue', 'orange', 'green', 'red']  # Colors for the pie chart
         labels = ['DOS', 'U2R', 'R2L', 'PROBE']

         # Ensure the order of attack_types_count matches the labels order
         attack_types_count = attack_types_count.reindex([label.lower() for label in labels])

         # Bar chart for number of attack types in each category
         plt.figure(figsize=(7, 5), dpi=100)
         plt.bar(labels, attack_types_count, color=colors_bar)
         plt.xlabel('Attack categories')
         plt.ylabel('Number of attack types in each category')
         plt.title('Attack categories and number of attack types in cyber security')
         plt.show()

         # Ensure the order of attack_numbers_sum matches the labels order
         attack_numbers_sum = attack_numbers_sum.reindex([label.lower() for label in labels])

         # Define the custom autopct function
         def absolute_value(val):
             return f'{val:.1f}'

         # Pie chart for number of attacks in each category
         explode = [0.1 if x == max(attack_numbers_sum) else 0 for x in attack_numbers_sum]

         plt.figure(figsize=(10, 10))
         plt.pie(attack_numbers_sum, labels=labels, colors=colors_pie, autopct=absolute_value, explode=explode, startangle=10)
         plt.title('Attack categories and number of attacks in cyber security')
         plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
         plt.show()
```

Attack categories and number of attack types in cyber security

Attack categories and number of attacks in cyber security