

|  |
| --- |
|  |
| Nexus Project |
| Data Analyst Projects:  Exploring Data Insights |
| KUMAR GAURAV | gauravt.nic@gmail.com |

|  |  |  |
| --- | --- | --- |
| Decorative |  |  |
| Introduction |  |
| **PROJECT 2: Seattle Weather Data Analysis**  **Author:**  Kumar Gaurav  **Date:**  30th May, 2024  **Purpose:**  The purpose of this project is to perform Exploratory Data Analysis (EDA) on the Seattle weather dataset to visualize key statistics and distributions, explore correlations, patterns, and trends, and document the findings using Power BI and Python.  **Data Description**  **Dataset Information:**   * **Name**: Seattle Weather Dataset * **Source**: The dataset is publicly available at Kaggle. * **Description**: The dataset contains historical weather data for Seattle. It includes various weather parameters such as temperature, precipitation, and wind speed..   **Attributes:**   1. **Date:** Date of the weather observation. 2. **Precipitation:** Amount of precipitation (rainfall or snowfall) in inches. 3. **Temperature (Max and Min):** Maximum and minimum temperature in Fahrenheit. 4. **Wind:** Wind speed in miles per hour. 5. **Weather:** Description of the weather conditions (e.g., sunny, cloudy, rainy).   **3. Data Loading and Cleaning**  **Loading Data:**  The dataset was loaded into the Python environment using the pandas library:  import pandas as pd  # Load the dataset  weather\_data = pd.read\_csv('seattle\_weather.csv')  **Data Cleaning:**  No missing values were found in the dataset. The data was found to be clean with no missing values. No outliers were handled as all data points were within expected ranges.  **4. Exploratory Data Analysis (EDA)**  **Statistical Summary:**  A statistical summary of the dataset was generated using **pandas**:  weather\_data.describe()  **Visualizations:**   * **Histogram**:   + The histogram for temperature (max) showed a normal distribution with a peak around 60 to 70 degrees Fahrenheit.   import matplotlib.pyplot as plt  weather\_data['temp\_max'].hist()  plt.title('Distribution of Max Temperature')  plt.xlabel('Temperature (F)')  plt.ylabel('Frequency')  plt.show()   * **Scatter Plot**:   + The scatter plot showed a positive correlation between maximum temperature and precipitation, with different weather conditions forming distinct clusters.   weather\_data.plot(kind='scatter', x='temp\_max', y='precipitation', c='weather', colormap='viridis')  plt.title('Max Temperature vs Precipitation')  plt.xlabel('Max Temperature (F)')  plt.ylabel('Precipitation (in)')  plt.show()   * **Box Plot**:   The box plot showed the distribution of wind speed for different weather conditions.  weather\_data.boxplot(column='wind', by='weather')  plt.title('Wind Speed by Weather Condition')  plt.xlabel('Weather Condition')  plt.ylabel('Wind Speed (mph)')  plt.show()   1. **Data Visualization with Power BI**   **Create Visualizations**:   * + Create histograms, scatter plots, and box plots for different weather parameters.   **Interactivity:**   * **Slicer**:   + Add slicers or filters for interactive data exploration.  1. **Insights and Conclusion**   **Key Findings:**   * Higher precipitation levels are observed during rainy weather conditions. * Maximum temperature shows a positive correlation with minimum temperature. * Wind speed varies across different weather conditions, with higher speeds during stormy weather.   **Conclusion**  The analysis provided valuable insights into the Seattle weather dataset. By visualizing the distributions and relationships between weather parameters, we can better understand the weather patterns in Seattle. The use of Python for EDA and Power BI for interactive visualizations complemented each other to enhance the overall understanding of the weather data.  **References**  UCI Machine Learning Repository: Seattle Weather Dataset |  |  |
|  |  |  |
|  |  |  |