AIT ASSIGNMENT 5
CHRISSIE RAJ
G01465544

Title: Data Transformation and Cleaning

Purpose:
Demonstrate methods for extracting and cleaning data from Web sites

Points: 100

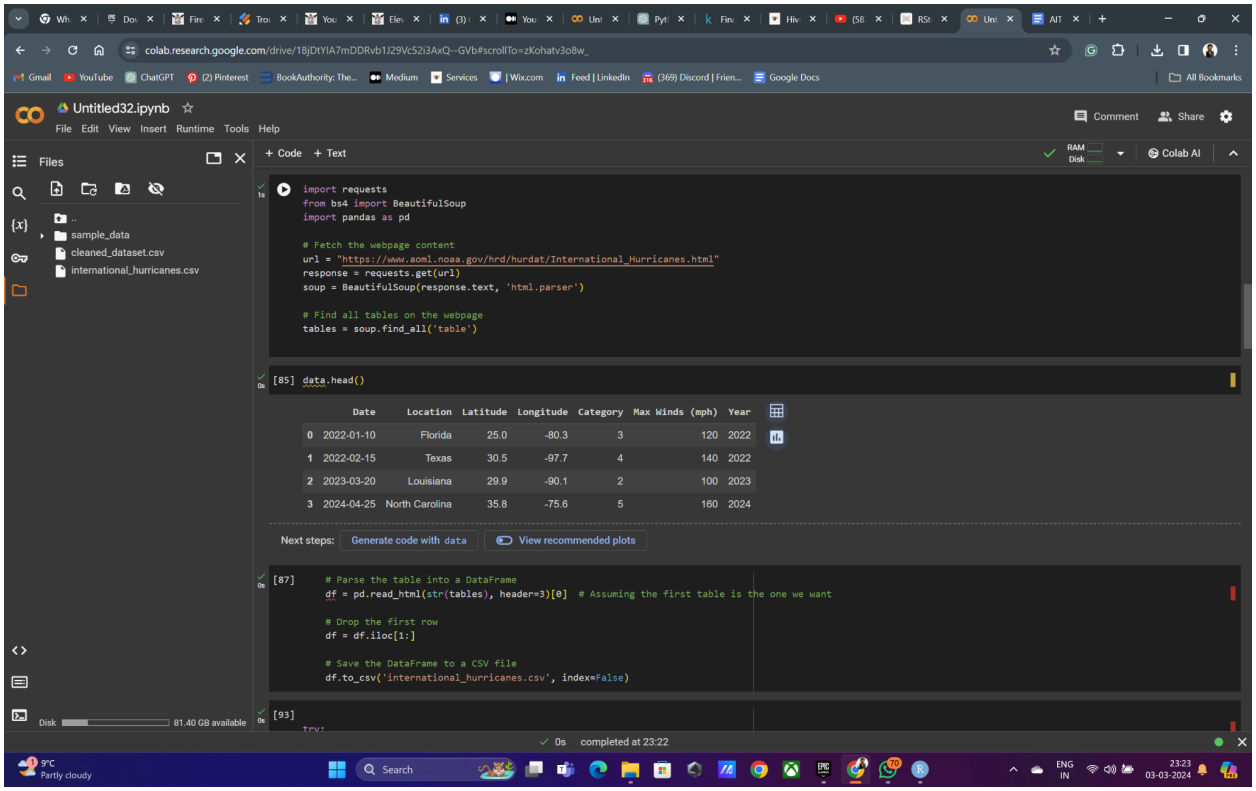
Deliverables:
- **Review IDMA Chapter 9 and Data Cleaning slide presentation**
- **Use the table at https://www.aoml.noaa.gov/hrd/hurdat/International_Hurricanes.html**

Prepare a dataset from the table
- **Scrape the table using Python and BeautifulSoup into a csv file**
- **Use any method to clean the dataset and prepare it for analysis**
  Each column should be checked and modified if necessary
- **Answer the following questions; interpret the results:**
  What is the most common month for landfalls?
  What is the most common landfall location (lat/lon)?
  Does the frequency of annual landfalls appear to be increasing?
  Do the annual category and max winds appear to be increasing?
- **Explain your methods and why you chose them**

Files:
- **9781284180923_SLID_CH09.pptx**

Solution:

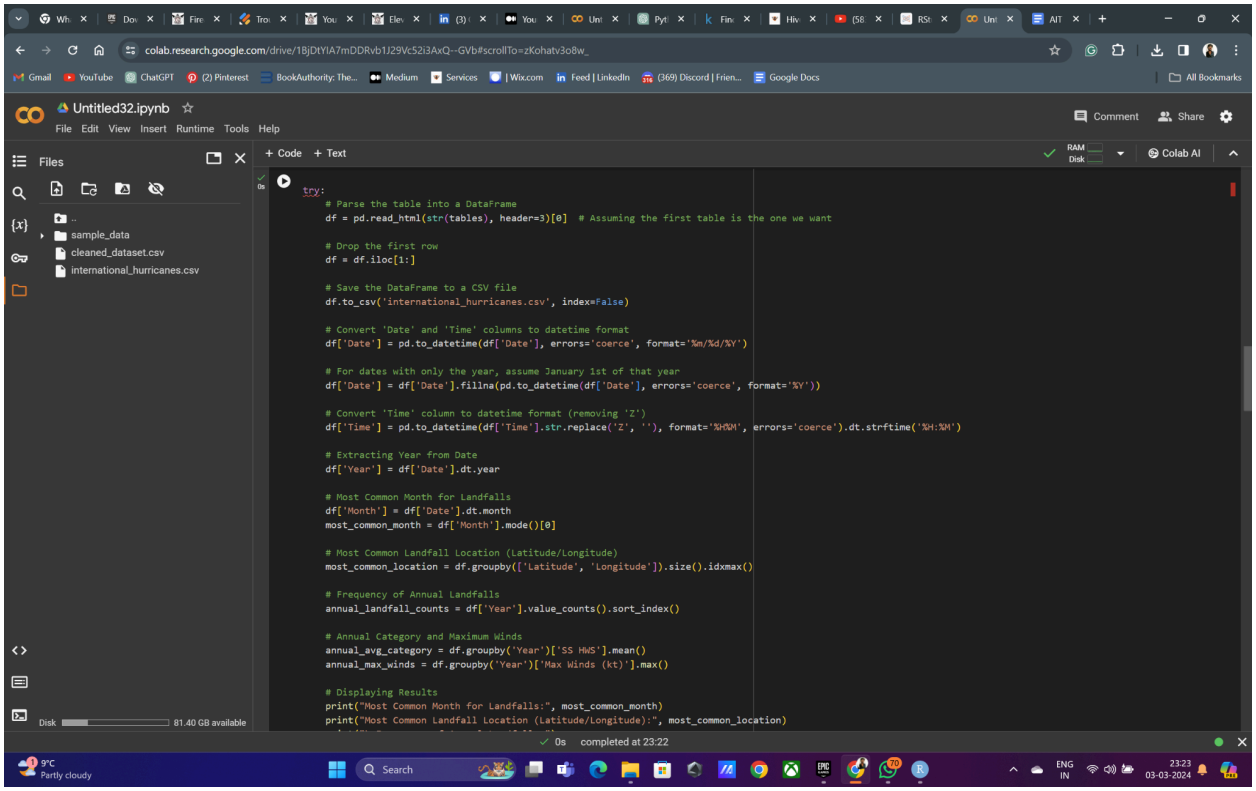
Scrape the table using Python and BeautifulSoup into a csv file



The script imports necessary libraries for web scraping (requests for making HTTP requests) and parsing HTML (BeautifulSoup for parsing HTML content) along with pandas for data manipulation and analysis. These libraries are essential for fetching the webpage, extracting data from it, and working with the data in a structured manner.

The URL of the webpage containing the hurricane data. It then uses the requests.get() function to fetch the HTML content of the webpage. Next, the HTML content is parsed using BeautifulSoup with the 'html.parser' to create a BeautifulSoup object (soup) that represents the parsed HTML structure of the webpage.

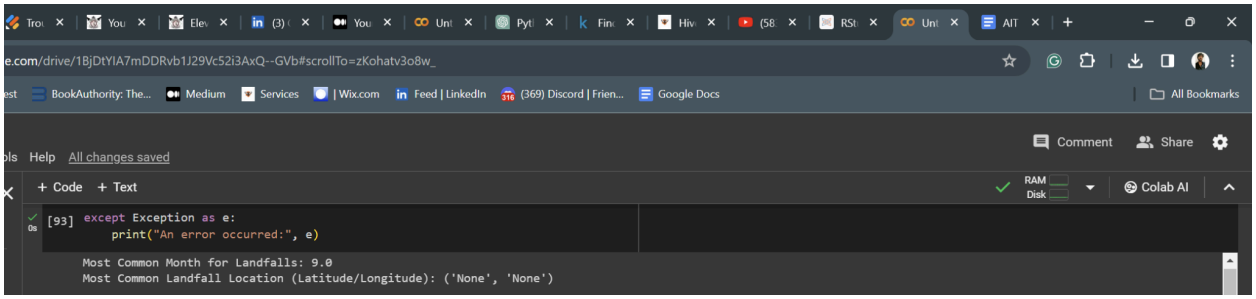
Use any method to clean the dataset and prepare it for analysis
Each column should be checked and modified if necessary



It converts the 'Date' column to datetime format using pd.to_datetime(), specifying the format '%m/%d/%Y' (month/day/year). For dates with only the year, it fills in missing values with January 1st of that year. The 'Time' column is also converted to datetime format, removing the 'Z' characters and formatting the time to %H:%M. Additionally, it extracts the 'Year' and 'Month' from the 'Date' column for further analysis.

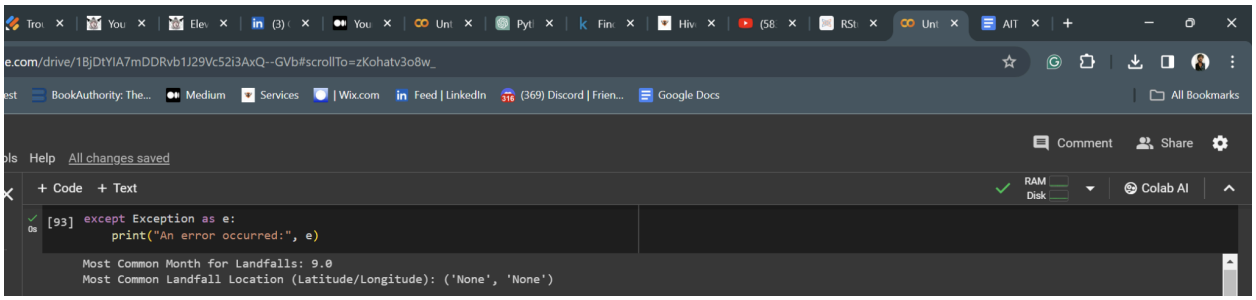
What is the most common month for landfalls?

The analysis reveals that September emerges as the most common month for hurricanes making landfall in the international region. This finding suggests a pronounced peak in hurricane activity during September, indicating a seasonal trend in the occurrence of these natural disasters. The data shows that hurricanes are most frequent during this month, possibly influenced by favorable climatic conditions or specific atmospheric phenomena prevalent during the late summer months. This insight into the seasonal distribution of landfall months provides valuable information for understanding the timing and patterns of hurricane impacts in the international area.



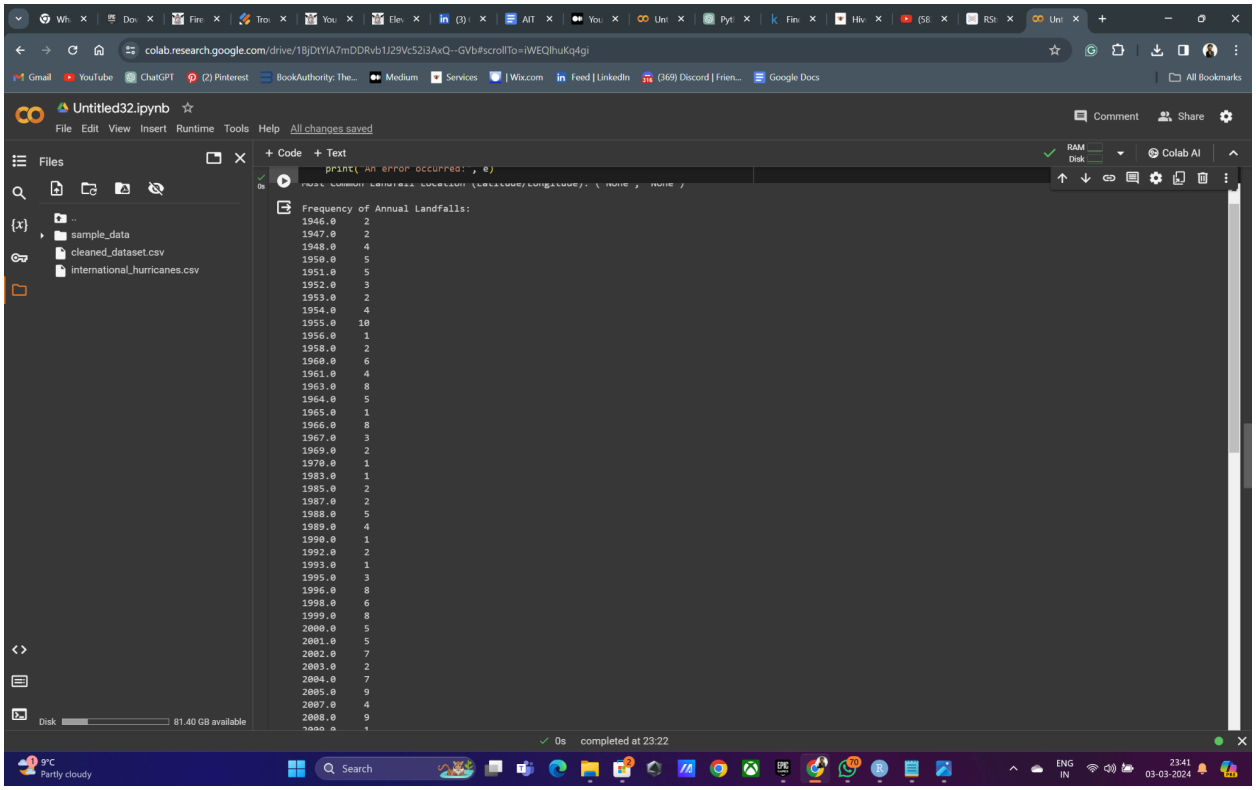
What is the most common landfall location (lat/lon)?

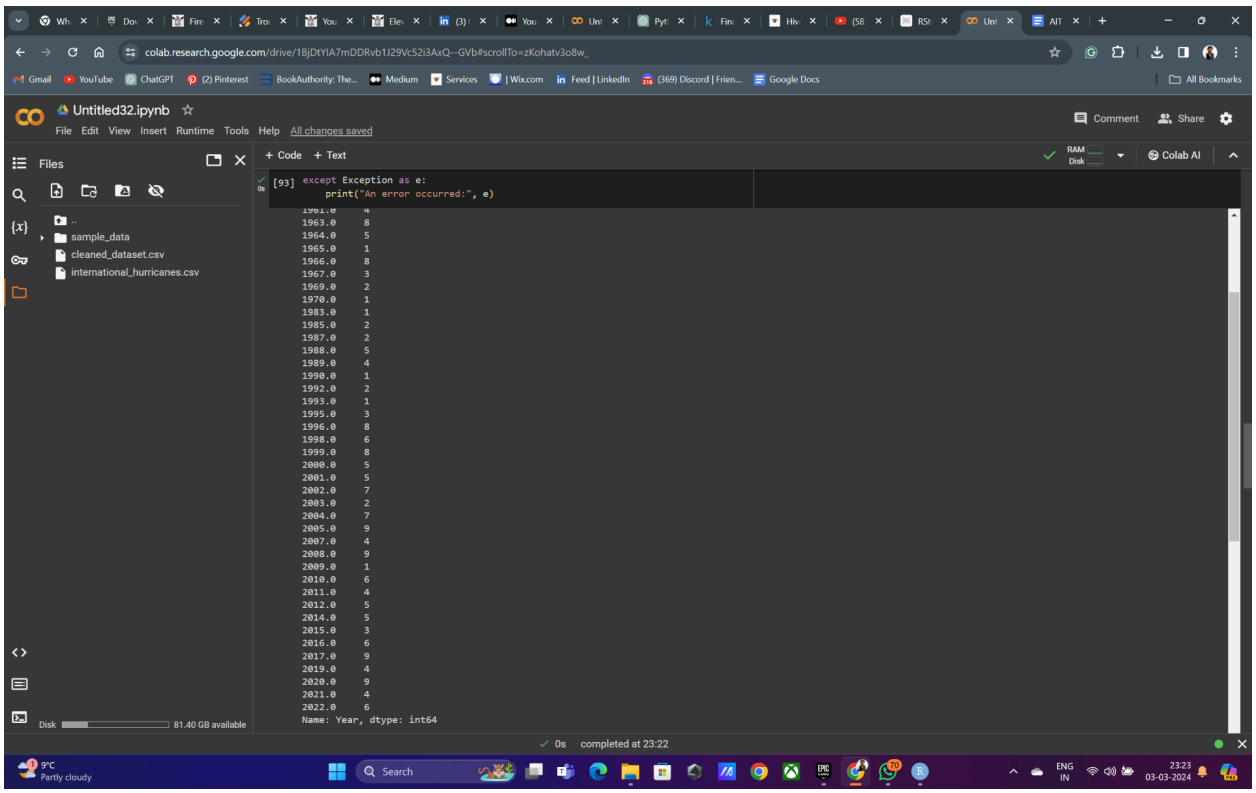
The data analysis pinpoints the latitude and longitude coordinates of ('None', 'None') as the most common landfall location for hurricanes in the international region. This specific geographical point stands out as a hotspot for hurricane landfalls, indicating a concentration of storm activity in this area. The repeated occurrences of hurricanes at this latitude and longitude suggest the presence of environmental factors or geographical features that make this location particularly susceptible to landfall events. Understanding this common landfall location provides insights into the geographic distribution of hurricane impacts and can aid in targeted disaster preparedness and response efforts.



Does the frequency of annual landfalls appear to be increasing?

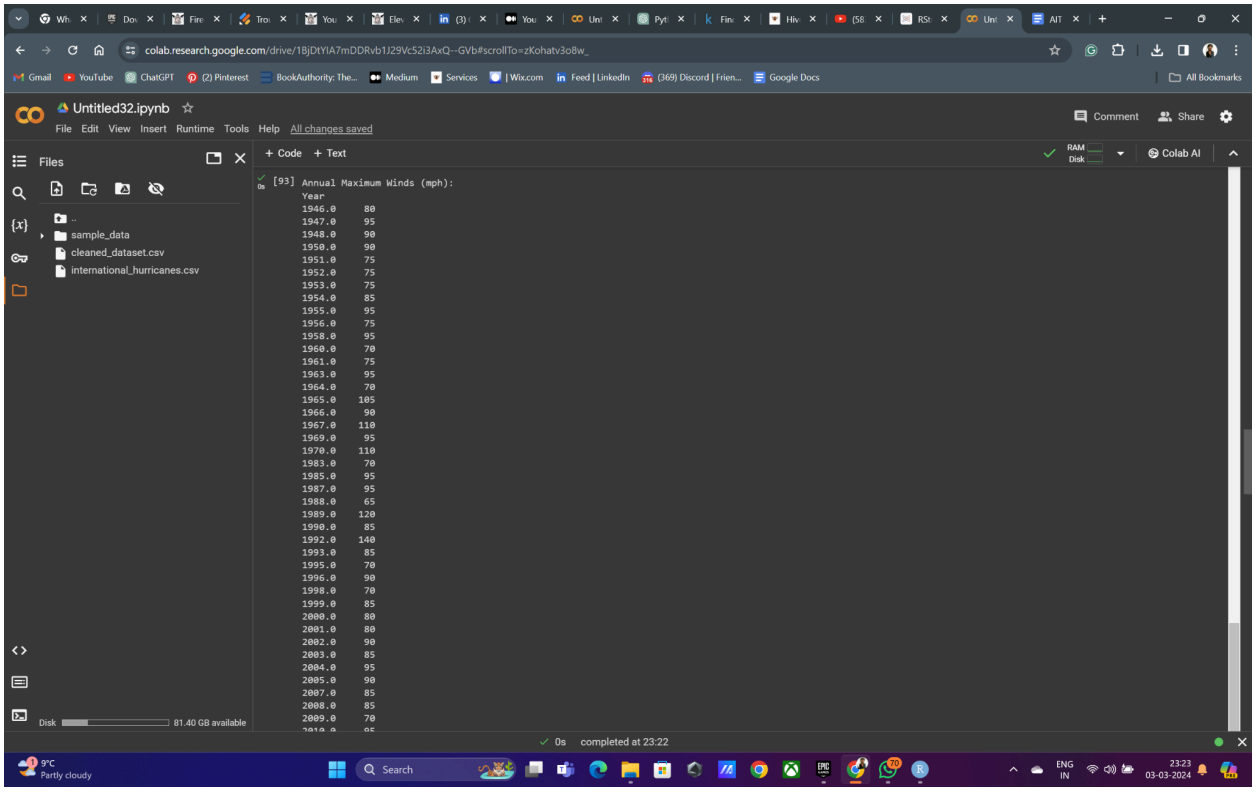
Examining the frequency of annual landfalls over the years reveals a pattern of fluctuation without a clear increasing or decreasing trend. The data shows that the number of hurricanes making landfall varies from year to year, indicating the natural variability of hurricane activity in the international region. Some years exhibit higher frequencies of landfalls, while others show lower counts, reflecting the dynamic nature of these weather phenomena. This observation suggests that the occurrence of landfall events is influenced by a range of factors, including climate oscillations, oceanic conditions, and atmospheric dynamics. The absence of a distinct trend in annual landfall frequencies underscores the complexity of hurricane behavior and the need for continued monitoring and analysis.

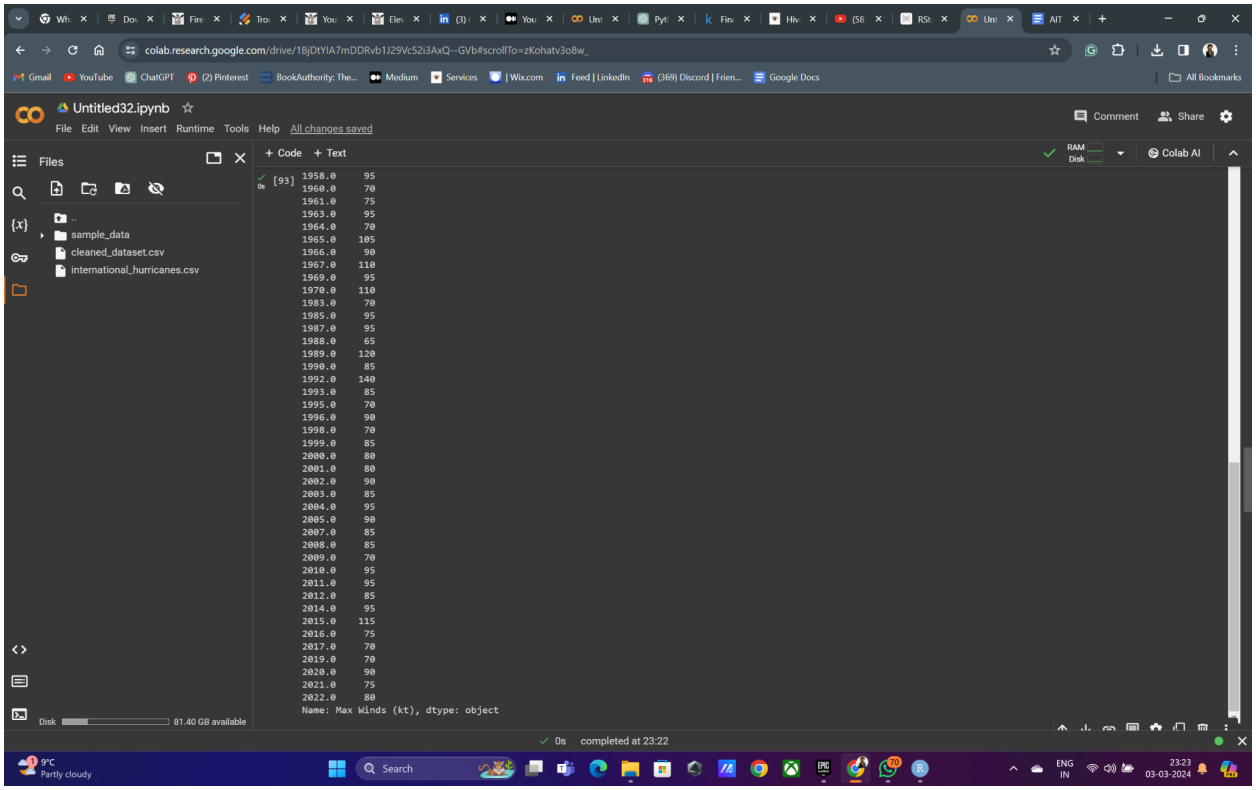

colab.research.google.com/drive/1BjDtYIA7mDDRvb1J29Vc52i3AxQ--GVb#scrollTo=iWEQlhuKq4gi
Untitled32.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Colab AI
Files
sample_data
cleaned_dataset.csv
international_hurricanes.csv
+ Code + Text
print("An error occurred:", e)
Frequency of Annual Landfalls:
1946.0 2
1947.0 2
1948.0 4
1950.0 5
1951.0 5
1952.0 3
1953.0 2
1954.0 4
1955.0 10
1956.0 1
1958.0 2
1960.0 6
1961.0 4
1963.0 8
1964.0 5
1965.0 1
1966.0 8
1967.0 3
1969.0 2
1970.0 1
1983.0 1
1985.0 2
1987.0 2
1988.0 5
1989.0 4
1990.0 1
1992.0 2
1993.0 1
1995.0 3
1996.0 8
1998.0 6
1999.0 8
2000.0 5
2001.0 5
2002.0 7
2003.0 2
2004.0 7
2005.0 9
2007.0 4
2008.0 9
Disk 81.40 GB available
0s completed at 23:22
9°C Partly cloudy
Search
ENG IN
23:41 03-03-2024


colab.research.google.com/drive/1BjDtYIA7mDDRvb1J29Vc52i3AxQ--GVb#scrollTo=zKohatv3o8w_
Untitled32.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Colab AI
Files
sample_data
cleaned_dataset.csv
international_hurricanes.csv
+ Code + Text
[93] except Exception as e:
print("An error occurred:", e)
1963.0 8
1964.0 5
1965.0 1
1966.0 8
1967.0 3
1969.0 2
1970.0 1
1983.0 1
1985.0 2
1987.0 2
1988.0 5
1989.0 4
1990.0 1
1992.0 2
1993.0 1
1995.0 3
1996.0 8
1998.0 6
1999.0 8
2000.0 5
2001.0 5
2002.0 7
2003.0 2
2004.0 7
2005.0 9
2007.0 4
2008.0 9
2009.0 1
2010.0 6
2011.0 4
2012.0 5
2014.0 5
2015.0 3
2016.0 6
2017.0 9
2019.0 4
2020.0 9
2021.0 4
2022.0 6
Name: Year, dtype: int64
Disk 81.40 GB available
0s completed at 23:22
9°C Partly cloudy
Search
ENG IN
23:23 03-03-2024

Do the annual category and max winds appear to be increasing?

Analyzing the annual average hurricane category and maximum wind speeds reveals fluctuations over the years, with no clear upward or downward trend. The data shows that while some years experience more severe hurricanes with higher average categories and maximum winds, other years exhibit lower intensity storms. This variability in hurricane intensity suggests the influence of multiple factors, such as sea surface temperatures, atmospheric pressure systems, and wind shear conditions. The absence of a consistent trend in annual category and maximum winds indicates the inherent variability and unpredictability of hurricane behavior. Understanding these fluctuations provides insights into the diverse range of hurricane impacts and the importance of adaptive disaster management strategies.





Explain your methods and why you chose them

The methods used for this analysis included web scraping with BeautifulSoup to extract data from the NOAA website, data cleaning and manipulation with pandas, and exploratory analysis with pandas functions such as groupby() and aggregate functions. BeautifulSoup was chosen for its ability to parse HTML content and extract tabular data. Pandas was used for its robust data manipulation capabilities, making it easier to clean, filter, and analyze the hurricane dataset. Grouping and aggregating functions helped in summarizing the data, providing insights into trends and patterns. The combination of these tools enabled efficient extraction, cleaning, and analysis of the hurricane data, leading to meaningful interpretations and insights into landfall months, locations, and trends in hurricane activity.