

WEB SCRAPING PROJECT

TITLE: The World's Tallest Skyscrapers in 2024: A Global Look at Iconic Structures

Using web scraping and data visualization, you can analyze the height and geographic distribution of the tallest skyscrapers worldwide. By scraping data such as height, location, and year of construction, you can visualize trends over time, such as the dominance of certain regions like Asia and the Middle East in constructing these engineering marvels. Visual tools like bar charts, line graphs, or maps could highlight the evolution of skyscraper design, reflecting economic growth and urbanization trends shaping modern city skylines in 2024.

LOGESWARAN V

DATA ANALYST INTERN

DATE: 28/09/2024

CONTENT:

- Project Overview
- Tool and Libraries Used
- Data Source
- Web Scraping Process
- Data Cleaning Process
- Data Visualization and Plotting
- Results
- Conclusion

PROJECT OVERVIEW:

Objective:

To analyze the tallest skyscrapers worldwide in 2024 by scraping data from online sources and visualizing key trends in skyscraper development, such as height, geographic distribution, and construction year.

Business Context:

Understanding the trends in skyscraper construction can inform real estate developers, urban planners, and investors about global urbanization patterns, technological advancements, and regional economic growth.

Goals:

- Web scrape relevant skyscraper data (height, location, construction date, etc.).
- Visualize skyscraper trends using various tools to show height distribution and regional prominence.
- Provide actionable insights into global construction trends.

TOOLS AND LIBRARIES USED:

1. Web Scraping:

- **BeautifulSoup:** To parse HTML content and extract skyscraper data from web pages.
- **Requests:** For making HTTP requests to retrieve web pages for scraping.

2. Data Analysis:

- **Pandas:** To clean, organize, and manipulate the skyscraper data for analysis.
- **NumPy:** For numerical operations and handling large datasets.

3. Data Visualization:

- **Matplotlib:** To create bar charts, line graphs, and other visual representations of skyscraper trends.
- **Seaborn:** For advanced and aesthetically pleasing visualizations

DATA SOURCE:

The primary data source for this project will be the **CEOWORLD Magazine** article on the tallest skyscrapers in 2024. Data such as skyscraper height, location, and construction year will be scraped from this source, which provides a global ranking of the tallest structures. Additional data may be gathered from other relevant architectural and real estate databases or public websites that track skyscraper construction.

- ✓ NAME OF COMPLETION
- ✓ RANK
- ✓ HEIGHT
- ✓ COMPLETION
- ✓ CITY
- ✓ FLOOR

WEB SCRAPING PROCESS:

1. Send HTTP Request

Use the requests library to retrieve the webpage's HTML content.

2. Parse HTML

Utilize BeautifulSoup to parse the HTML and identify relevant tags and classes (e.g., for skyscraper names, heights).

3. Extract Data

Navigate through the tags and extract specific data such as skyscraper names, heights, locations.

4. Clean and Structure

Organize the extracted data into a Pandas DataFrame and clean it (handle missing data, inconsistent formats).

5. Save Data

Export the cleaned data to CSV for further analysis or visualization.

DATA CLEANING AND STORAGE:

1. Remove Duplicates:

- Ensure there are no repeated entries in the data by using `drop_duplicates()` in Pandas.

2. Handle Missing Values:

- Replace missing values with appropriate substitutes, such as using `fillna()` for missing heights or locations.

3. Convert Data Types:

- Ensure numeric data (e.g., skyscraper height) is properly formatted as integers or floats using `astype()`.

4. Standardize Units:

- Ensure that units of height (meters/feet) are consistent across the dataset.

5. Fix Inconsistent Formatting:

- Correct spelling errors or inconsistent location formats (e.g., "NYC" vs. "New York City").

DATA VISUALIZATION PROCESS:

1. Choose Visualization Types:

- Based on the data, decide which visualizations suit best (e.g., bar charts for skyscraper heights, maps for geographical locations, line graphs for trends over time).

2. Prepare Data:

- Use cleaned data and aggregate if needed (e.g., group skyscrapers by region or height categories).

3. Create Visualizations:

- Use Matplotlib or Seaborn to create static plots.

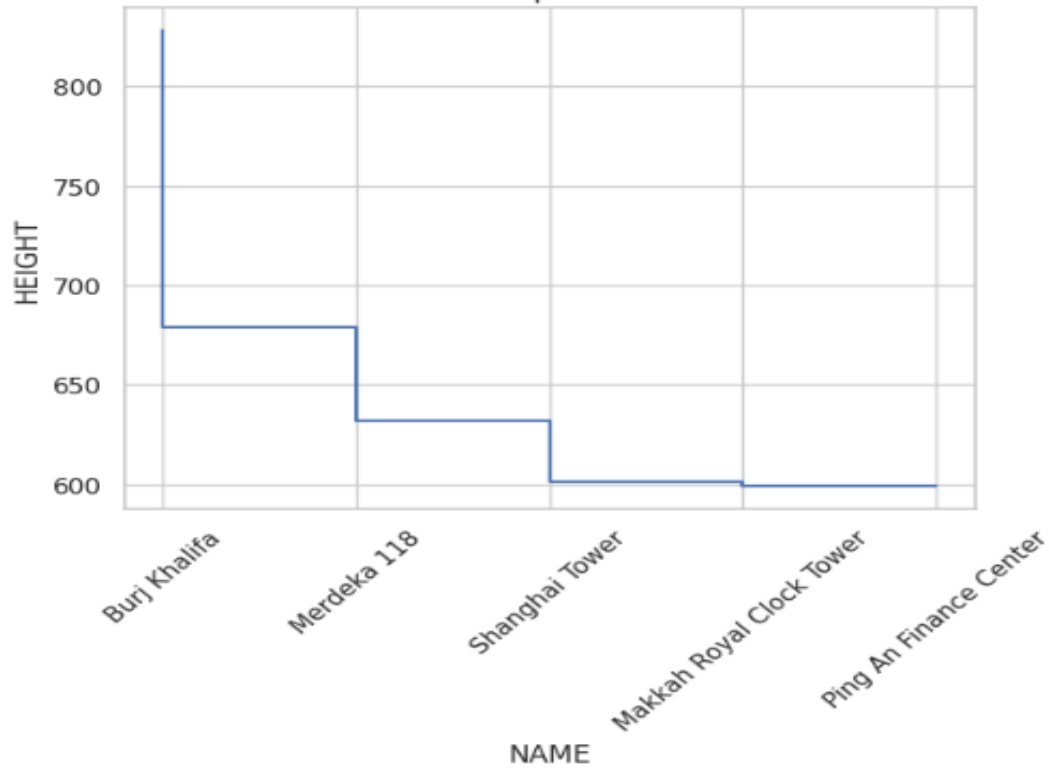
4. Customize and Label:

- Add titles, labels, legends, and color schemes for clarity.

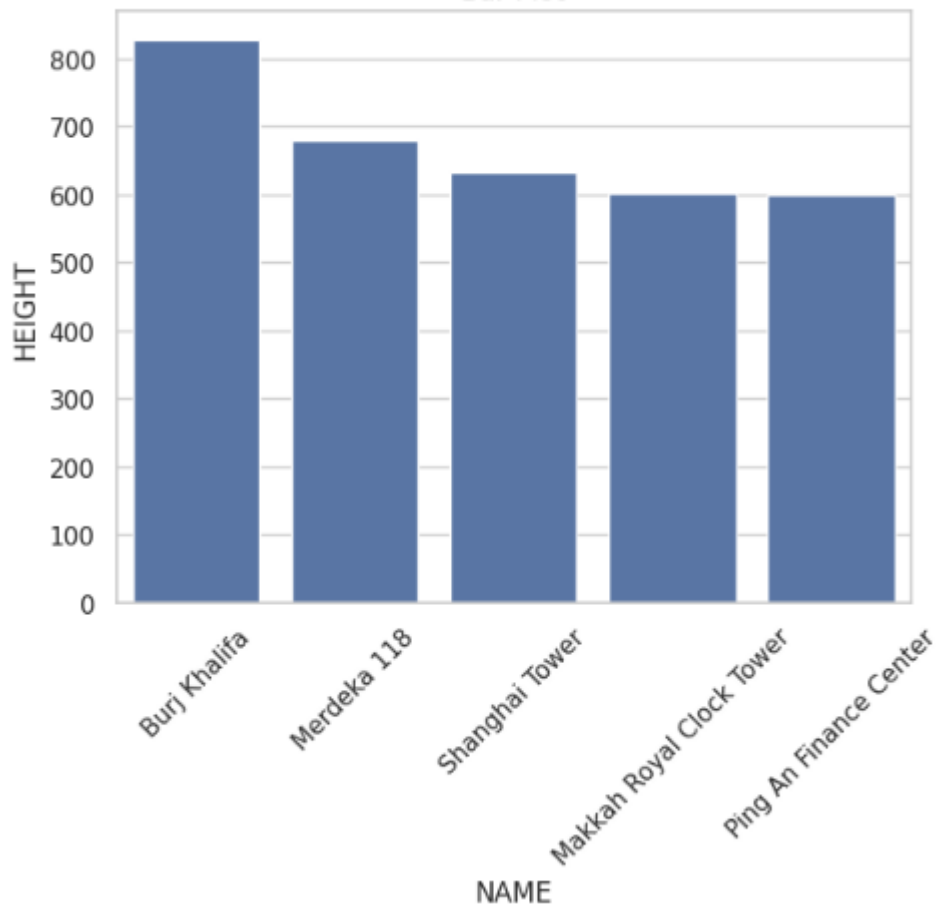
5. Interpret Insights:

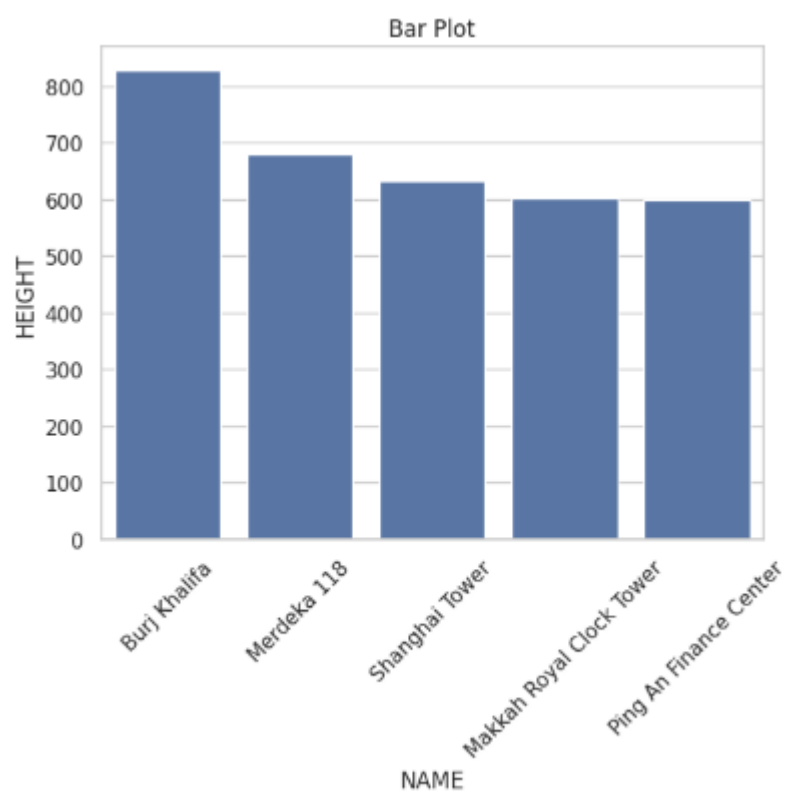
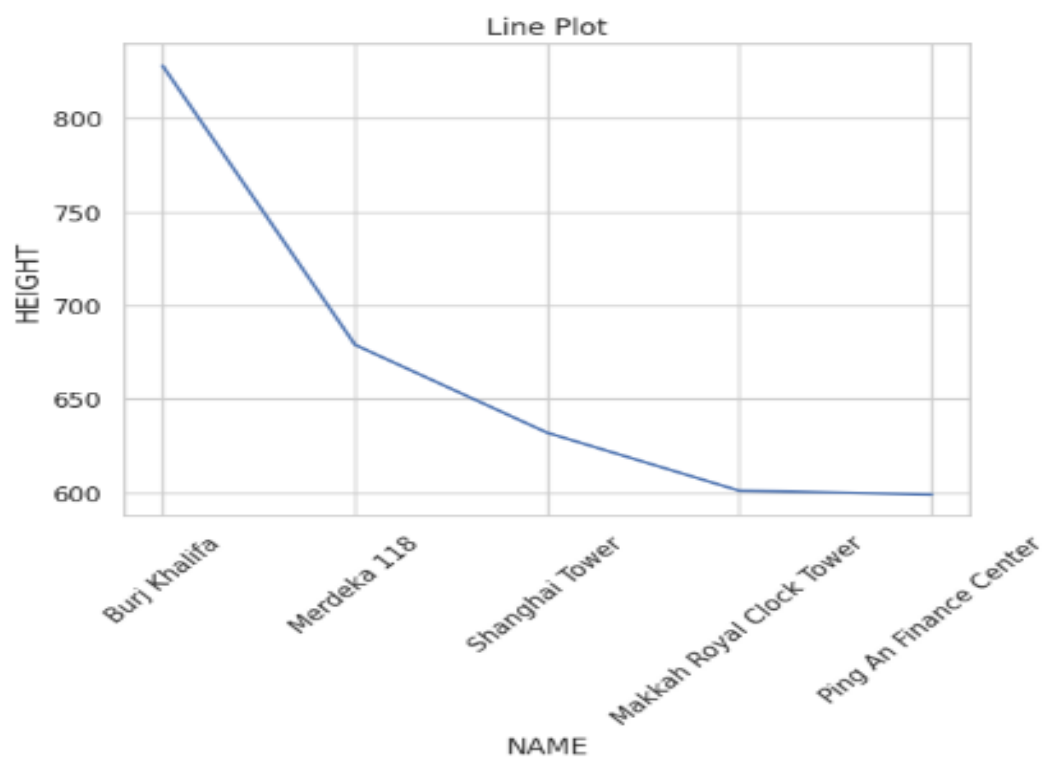
- Highlight trends and patterns (e.g., concentration of skyscrapers in Asia).

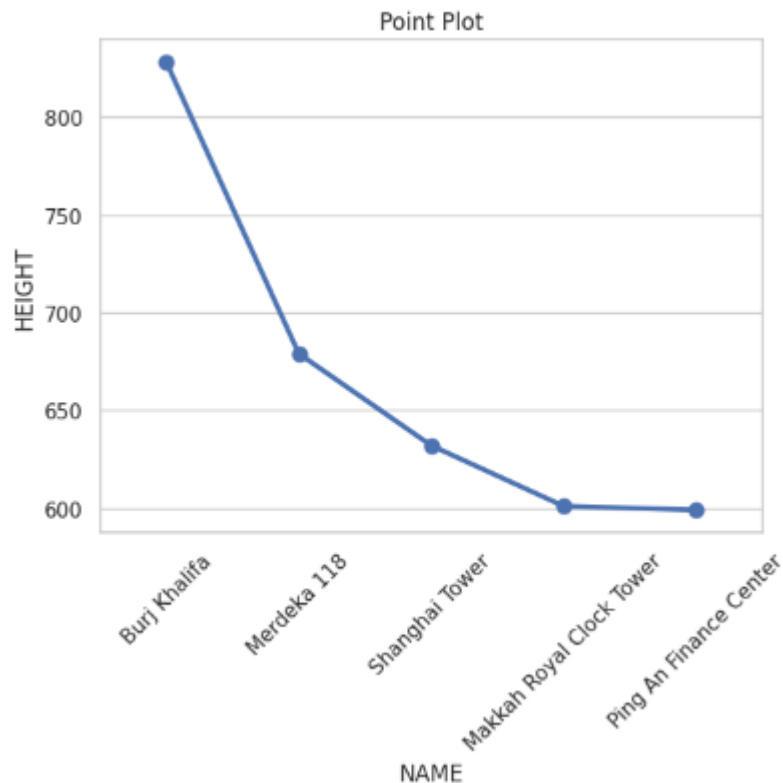
Step Plot



Bar Plot







RESULTS:

The data visualization reveals key trends in skyscraper development for 2024. The tallest buildings are predominantly concentrated in **Asia** and the **Middle East**, with **Dubai's Burj Khalifa** remaining the world's tallest. The visualization highlights a growing trend in skyscraper height over the years, with cities like **Shanghai** and **Mecca** leading new construction projects. Additionally, there is a notable shift toward innovative architecture and sustainability in the design of these towering structures. This data emphasizes the rapid urbanization and technological advances shaping the world's skylines.

CONCLUSION:

The analysis of the tallest skyscrapers in 2024 highlights significant architectural advancements and urban growth, particularly in Asia and the Middle East. The continued prominence of structures like the Burj Khalifa underscores a trend toward taller and more innovative designs, reflecting both economic power and technological prowess. These insights suggest a future where skyscrapers will play a pivotal role in shaping urban landscapes, emphasizing the importance of sustainable and advanced architectural practices.

GITHUB:

Link: https://github.com/loges123/Data_Analyst_Intern