Wikipedia Data Project – Web Scraping and Data Analysis

Project Overview

Select a Wikipedia page containing structured data (such as tables, rankings, or statistics) from the list provided below or directly from Wikipedia. Your goal is to extract and analyze data to generate meaningful insights and Key Performance Indicators (KPIs).

Data Selection

Choose one Wikipedia page that includes structured information such as:

- Statistical data

- Country-wise comparisons

- Rankings or indicators

- Any other form of tabular or structured data

You may use any Wikipedia link from the provided list or find one of your own.

NOTE: Every student has to choose unique dataset.

Datasets

Economics & Finance

1. https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal) - Countries ranked by nominal GDP in US dollars

2. https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP) - Countries ranked by GDP at Purchasing Power Parity

3. https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita - Countries by nominal GDP per person

4. https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita - Countries by PPP-adjusted GDP per person

5. https://en.wikipedia.org/wiki/List_of_countries_by_real_GDP_growth_rate - Countries by annual real GDP growth percentage

6. https://en.wikipedia.org/wiki/List_of_countries_by_tax_revenue - Countries with tax revenue as percentage of GDP

7. https://en.wikipedia.org/wiki/List_of_countries_by_average_wage - Average annual wages by country in USD PPP

8. https://en.wikipedia.org/wiki/List_of_countries_by_unemployment_rate - Countries by unemployment rate percentage

9. https://en.wikipedia.org/wiki/List_of_countries_by_inflation_rate - Countries by annual inflation rate

10. https://en.wikipedia.org/wiki/List_of_countries_by_minimum_wage - Minimum wage levels by country

11. https://en.wikipedia.org/wiki/List_of_countries_by_imports - Countries by total imports value

12. https://en.wikipedia.org/wiki/List_of_countries_by_exports - Countries by total exports of goods and services

13. https://en.wikipedia.org/wiki/List_of_countries_by_government_debt - Government debt as percentage of GDP by country

14. https://en.wikipedia.org/wiki/List_of_countries_by_external_debt - External debt levels by country

15. https://en.wikipedia.org/wiki/List_of_largest_companies_by_revenue - World's largest companies by revenue

Trade & Industry

16. https://en.wikipedia.org/wiki/List_of_countries_by_motor_vehicle_production - Countries by vehicle production numbers

17. https://en.wikipedia.org/wiki/List_of_countries_by_vehicle_exports - Countries by automotive exports value

18. https://en.wikipedia.org/wiki/List_of_countries_by_steel_production - Steel production by country in tonnes

19. https://en.wikipedia.org/wiki/List_of_countries_by_oil_production - Crude oil production by country

20. https://en.wikipedia.org/wiki/List_of_countries_by_oil_exports - Countries by crude oil export volume

21. https://en.wikipedia.org/wiki/List_of_countries_by_coal_production - Major coal-producing countries

22. https://en.wikipedia.org/wiki/List_of_countries_by_coffee_production - Coffee production by country

23. https://en.wikipedia.org/wiki/List_of_countries_by_wheat_production - Wheat production statistics by country

24. https://en.wikipedia.org/wiki/List_of_countries_by_rice_production - Rice production by country

25. https://en.wikipedia.org/wiki/List_of_countries_by_milk_production - Milk production statistics by country

Demographics & Social

26. https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population - Countries ranked by population size

27. https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population_density - Countries by population per square kilometer

28. https://en.wikipedia.org/wiki/List_of_countries_by_median_age - Countries by median population age

29. https://en.wikipedia.org/wiki/List_of_countries_by_fertility_rate - Total fertility rate by country

30. https://en.wikipedia.org/wiki/List_of_countries_by_population_growth_rate - Annual population growth rates by country

31. https://en.wikipedia.org/wiki/List_of_countries_by_net_migration_rate - Net migration per 1,000 people by country

32. https://en.wikipedia.org/wiki/List_of_countries_by_Human_Development_Index - Countries ranked by HDI

33. https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy - Life expectancy statistics by country

34. https://en.wikipedia.org/wiki/List_of_countries_by_obesity_rate - Adult obesity rates by country

35. https://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate - Adult literacy rates by country

36. https://en.wikipedia.org/wiki/List_of_countries_by_infant_mortality_rate - Infant mortality rates by country

37. https://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate - Suicide rates by country

Energy & Environment

38. https://en.wikipedia.org/wiki/List_of_countries_by_electricity_consumption - Electricity consumption by country

39. https://en.wikipedia.org/wiki/List_of_countries_by_electricity_production - Electricity production by country

40. https://en.wikipedia.org/wiki/List_of_countries_by_carbon_dioxide_emissions - Total CO2 emissions by country

41. https://en.wikipedia.org/wiki/List_of_countries_by_carbon_dioxide_emissions_per_capita - CO2 emissions per person by country

42. https://en.wikipedia.org/wiki/List_of_countries_by_greenhouse_gas_emissions - Total GHG emissions by country

43. https://en.wikipedia.org/wiki/List_of_countries_by_renewable_electricity_production - Renewable electricity production by country

44. https://en.wikipedia.org/wiki/List_of_countries_by_forest_area - Forest coverage by country

45. https://en.wikipedia.org/wiki/List_of_countries_by_natural_gas_production - Natural gas production by country

46. https://en.wikipedia.org/wiki/Nuclear_power_by_country - Nuclear power plants and generation by country

47. https://en.wikipedia.org/wiki/List_of_countries_by_uranium_reserves - Uranium reserves by country

48. https://en.wikipedia.org/wiki/List_of_countries_by_arable_land_density - Arable land statistics by country

49. https://en.wikipedia.org/wiki/List_of_countries_by_renewable_energy_consumption - Renewable energy consumption by country

Technology & Internet

50. https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users - Internet users by country

51. https://en.wikipedia.org/wiki/List_of_countries_by_Internet_connection_speeds - Internet connection speeds by country

52. https://en.wikipedia.org/wiki/List_of_most-visited_websites - Top global websites by traffic

53. https://en.wikipedia.org/wiki/World_Intellectual_Property_Indicators - Patent and trademark filings by country

54. https://en.wikipedia.org/wiki/List_of_countries_by_research_and_development_spending - R&D spending by country

55. https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration - Smartphone penetration rates by country

56. https://en.wikipedia.org/wiki/List_of_major_stock_exchanges - Stock exchanges by market capitalization

57. https://en.wikipedia.org/wiki/List_of_countries_by_number_of_scientific_and_technical_journal_articles - Scientific publications by country

Sports & Entertainment

58. https://en.wikipedia.org/wiki/All-time_Olympic_Games_medal_table - All-time Olympic medal counts by country

59. https://en.wikipedia.org/wiki/List_of_FIFA_World_Cup_finals - FIFA World Cup final matches history

60. https://en.wikipedia.org/wiki/List_of_Grand_Slam_singles_champions_by_country - Tennis Grand Slam winners by country

61. https://en.wikipedia.org/wiki/List_of_best-selling_video_games - Best-selling video games of all time

62. https://en.wikipedia.org/wiki/List_of_highest-grossing_films - Highest-grossing films worldwide

63. https://en.wikipedia.org/wiki/List_of_FIFA_World_Cup_winners - FIFA World Cup winners by year

64. https://en.wikipedia.org/wiki/List_of_Olympic_Games_host_cities - Olympic Games host cities

65. https://en.wikipedia.org/wiki/List_of_UEFA_Champions_League_finals - UEFA Champions League finals history

66. https://en.wikipedia.org/wiki/List_of_ICC_Cricket_World_Cup_finals - Cricket World Cup finals

67. https://en.wikipedia.org/wiki/List_of_best-selling_music_artists - Best-selling music artists

68. https://en.wikipedia.org/wiki/List_of_highest-grossing_media_franchises - Highest-grossing media franchises

69. https://en.wikipedia.org/wiki/List_of_most-subscribed_YouTube_channels - Most-subscribed YouTube channels

70. https://en.wikipedia.org/wiki/List_of_most-followed_Instagram_accounts - Most-followed Instagram accounts

Geography & Infrastructure

71. https://en.wikipedia.org/wiki/List_of_tallest_buildings_by_country - Tallest buildings by country

72. https://en.wikipedia.org/wiki/List_of_countries_by_highest_point - Highest elevation points by country

73. https://en.wikipedia.org/wiki/List_of_highest_mountains_on_Earth - Highest mountains globally

74. https://en.wikipedia.org/wiki/List_of_river_systems_by_length - Longest river systems worldwide

75. https://en.wikipedia.org/wiki/List_of_largest_lakes - Largest lakes in the world

76. https://en.wikipedia.org/wiki/List_of_stadiums_by_capacity - Stadiums by seating capacity

77. https://en.wikipedia.org/wiki/List_of_tallest_bridges - Tallest bridges in the world

78. https://en.wikipedia.org/wiki/List_of_longest_tunnels - Longest tunnels worldwide

79. https://en.wikipedia.org/wiki/List_of_largest_dams - Largest dams in the world

80. https://en.wikipedia.org/wiki/List_of_European_cities_by_population_within_city_limits - European cities by population

81. https://en.wikipedia.org/wiki/List_of_metro_systems - Metro systems worldwide

82. https://en.wikipedia.org/wiki/List_of_largest_ports - Largest ports by cargo volume

Transportation

83. https://en.wikipedia.org/wiki/List_of_busiest_airports_by_passenger_traffic - Busiest airports by passenger traffic

84. https://en.wikipedia.org/wiki/List_of_busiest_airports_by_international_passenger_traffic - Busiest international airports

85. https://en.wikipedia.org/wiki/List_of_countries_by_rail_transport_network_size - Rail network size by country

86. https://en.wikipedia.org/wiki/List_of_countries_by_road_network_size - Road network size by country

87. https://en.wikipedia.org/wiki/List_of_busiest_container_ports - Busiest container ports

88. https://en.wikipedia.org/wiki/List_of_airlines_by_fleet_size - Airlines by fleet size

89. https://en.wikipedia.org/wiki/World_Tourism_rankings - International tourist arrivals by country

90. https://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate - Road traffic deaths by country

Natural Disasters

91. https://en.wikipedia.org/wiki/List_of_historical_earthquakes - Major historical earthquakes

92. https://en.wikipedia.org/wiki/Lists_of_earthquakes - Earthquake lists by region and time

93. https://en.wikipedia.org/wiki/List_of_volcanic_eruptions_by_death_toll - Deadliest volcanic eruptions

94. https://en.wikipedia.org/wiki/List_of_floods - Notable floods in history

95. https://en.wikipedia.org/wiki/List_of_tropical_cyclones - Tropical cyclones by year

96. https://en.wikipedia.org/wiki/List_of_deadliest_natural_disasters - Deadliest natural disasters

Military & Defense

97. https://en.wikipedia.org/wiki/List_of_countries_with_highest_military_expenditures - Military spending by country

98. https://en.wikipedia.org/wiki/List_of_countries_by_military_expenditure_per_capita - Military spending per capita

99. https://en.wikipedia.org/wiki/List_of_countries_by_number_of_military_and_paramilitary_personnel - Military personnel by country

100. https://en.wikipedia.org/wiki/List_of_countries_by_number_of_tanks - Military tanks by country

101. https://en.wikipedia.org/wiki/List_of_countries_by_number_of_aircraft - Military aircraft by country

Health

102.　　　https://en.wikipedia.org/wiki/List_of_countries_by_hospital_beds - Hospital beds per capita by country

103.　　　https://en.wikipedia.org/wiki/List_of_countries_by_body_mass_index - Average BMI by country

104.　　　https://en.wikipedia.org/wiki/List_of_countries_by_alcohol_consumption_per_capita - Alcohol consumption per capita

105.　　　https://en.wikipedia.org/wiki/List_of_countries_by_HIV/AIDS_adult_prevalence_rate - HIV/AIDS prevalence by country

106.　　　https://en.wikipedia.org/wiki/List_of_countries_by_intentional_homicide_rate - Homicide rates by country

Education & Culture

107.　　　https://en.wikipedia.org/wiki/List_of_largest_universities_and_university_networks_by_enrollment - Largest universities by enrollment

108.　　　https://en.wikipedia.org/wiki/List_of_countries_by_tertiary_education_attainment - Tertiary education attainment by country

109.　　　https://en.wikipedia.org/wiki/List_of_Nobel_laureates_by_country - Nobel Prize winners by country

110.　　　https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers - Languages by native speakers

111.　　　https://en.wikipedia.org/wiki/List_of_famines - Historical famines worldwide

Project Deliverables

1. Key Performance Indicators (KPIs)

- Develop **10–15 KPIs** based on your dataset.

- Each KPI should include:

  o A **clear definition**

  o **Visualizations** (charts or graphs) with properly labeled axes and descriptive titles

  o A short explanation of insights derived from the data

2. Insights and Storytelling Presentation

- Create a **PowerPoint presentation (minimum 10 slides)** that includes:

  o Key insights supported by visualizations and analysis

  o Screenshots of important findings

  o A clear and logical **storyline** that explains your analytical process

3. Code and Documentation

- Submit a **Jupyter Notebook** or **Python script** containing:

    o   Well-commented code for web scraping and analysis

    o   A **README.md** file explaining the project, tools, and steps

    o   A **requirements.txt** or **pyproject.toml** file listing all dependencies

---

Submission Details

- **Deadline:** 16 November 2025, by **18:00**

- **Submission Method:** Microsoft Teams (via personal chat)

- **Files to Submit:**

    1. **Notebook/Python Script** – include detailed comments and documentation

    2. **Presentation (PPT)** – summarize insights, visualizations, and storyline

Note: Zip the all the files and share in teams.

You will present your project during the **afternoon session on 18 November 2025** in class.

Evaluation Criteria

Your project will be assessed based on:

- Quality and relevance of KPIs

- Depth and complexity of data analysis

- Clarity and structure of insights and storytelling

- Presentation quality and overall professionalism