**Individual Project Documentation**

**Nandika Jhunjhunwala**

Start date: September 1st.

Time Spent: 3 hours

**Introduction**

With two datasets from Kaggle, one on AI Development Index for countries, and other on the AI threat to jobs index, I plan on creating impactful visualizations to depict the recent Artificial Intelligence revolution we’re experiencing in terms of development and progress. On the other hand, I also plan on visualizing how the rise of Artificial Intelligence threatens existing jobs.

**Datasets:**

**AI Global Index**

<https://www.kaggle.com/datasets/katerynameleshenko/ai-index>

CSV file contains 62 rows and 13 columns. Five columns are categorical, and the others are numerical in float format.

**SOURCES**

The data was collected in open sources: <https://www.tortoisemedia.com/intelligence/global-ai/>, <https://www.worldbank.org/en/home>, <https://ourworldindata.org/>

**COLLECTION METHODOLOGY**

The data was collected with web scrapping

A screenshot of a computer screen

Description automatically generated

The variables in this dataset are:

|  |  |
| --- | --- |
| Country | The country of interest |
| Talent | The AI Talent index on a scale of 0 – 100 |
| Infrastructure | The AI Infrastructure on a scale of 0 – 100 |
| Operating Environment | The Operating Environment for AI on a scale of 0 -100 |
| Research | Amount of AI Research done on a scale of 0 -100 |
| Region | The region the country belongs in. Classifications include “Americas”, “Europe”, “Asia Pacific”, “Middle East”, and “Africa” |
| Development | The country’s AI Development Score |
| Government Strategy | The country’s government strategy’s impact on AI, on a scale of 0 – 100 |
| Commercial | The country’s commercialisation of AI, on a scale of 0 -100 |
| Total Score | A cumulative AI Development Index score based on the previous scores, on a scale of 0-100 |
| Cluster | Classification of countries according to their AI experience as “Waking up”, “Rising stars”, “Nascent”, “Power players”, “Traditional Champions” |
| Income group | Average Income classification as “High”, “Upper Middle” or “Lower Middle”, |
| Political Regime | Classification of government regime as “Electoral autocracy”, “Closed autocracy”, “Liberal democracy”, “Electoral Democracy” |

A graph of different countries/regions

Description automatically generated with medium confidence

**From Data Entry to CEO: The AI Job Threat Index**

<https://www.kaggle.com/datasets/manavgupta92/from-data-entry-to-ceo-the-ai-job-threat-index>

This CSV file contains 4706 rows and 6 columns. 3 columns are strings, 2 are integers and 1 is a float.

Contained within this file is a comprehensive dataset examining the interplay between job roles and the pervasive influence of artificial intelligence (AI). Each entry meticulously categorizes job titles, shedding light on AI's potential impact, the workload distribution between human tasks and AI models, and the domain to which each job belongs. This database contains 4706 unique values or jobs.

**SOURCES**

The data has been sourced primarily from leading AI tools websites that provide insights into job analytics, AI implementations in the corporate world, and the broader impact of AI on the job market. By combining data from multiple trusted platforms, the dataset offers a holistic view, capturing both the depth and breadth of the evolving job landscape in the AI era.

**COLLECTION METHODOLOGY**

The "Job Threat Index" dataset was meticulously assembled using a combination of web scraping techniques and manual curation. Advanced scraping tools were employed to extract relevant data from a selection of prominent AI tools websites known for their insights into job analytics and AI's workplace impact. These tools navigated the intricate web structures, extracting job titles, AI impacts, task counts, and other relevant metrics. Post-extraction, the data was subjected to rigorous manual reviews, ensuring any anomalies, outliers, or inconsistencies were addressed. This dual approach, blending automated data harvesting with human oversight, ensured the dataset's comprehensiveness, accuracy, and relevance to the overarching theme of AI's influence on the modern job landscape.

A close-up of numbers

Description automatically generated

|  |  |
| --- | --- |
| Job\_Titles | The job title |
| AI Impact | The impact of AI threat on the job, on a percent scale from 0 – 100 |
| Tasks | The number of tasks related to the job that can be automated with AI, ranging from 1 to 1387 |
| AI\_Models | The number of AI models that are used to substitute work associated with that job, ranging from 0 to 5666 |
| AI\_Workload\_Ratio | A computed ratio representing the workload distribution between tasks and AI models. |
| Domain | The broader category or industry to which the job title belongs. |

**9th September:**

**Time Spent: 3 hours**

I am designinfg this app keeping in mind a wide audience of public so that they can be more educated about the Artificial Intelligence revolution. Started learning RShiny in class, spent time writing an RShiny application for my Individual Project. I also cleaned my data for the project. I spent some time cleaning my datasets to suit my visualization needs. For the AI threat dataset, I created two different data sets, I filtered so that I could depict jobs that were most impacted by AI.

**9th September**:

**Time Spent: 2 hours**

Exploring different visualisations, trying to render previously made graphs onto RShiny application. I also implemented a Navigation bar with each page containing a different visualisation.

Visualization 1: AI Development scores for countries

The application lets you select the variable you want visualised for the countries, and creates a bar plot. This is helpful for a comparative analysis of the global progress with Artificial I Intelligence.

Visualization 2: Barplot for AI Threat to most impacted jobs’

The second visualization also lets you select the variable you want to visualise and creates barplots for the selected occupations for a comparative analysis.

**16th September**:

**Time Spent: 5 hours**

Visualzation 3: Creating an interactive world map visualization so get a better understanding of development indices for AI on a global scale.

This was tricky to create since polygons for the leaflet library was difficult to merge with my dataset, so this visualization took a long time to complete

**23rd September**:

**Time Spent: 3 hours**

Visualzation 4: Creating a word cloud visualization to see which industries were most affected by AI.

I used the word cloud package and generated a corpus of texts and then counted the frequency of which industries appear the most in my dataset, and used the word frequency table to generate a word cloud.

**1st October**

**Time Spent: 4 hours**

I was trying to put all the visualizations together in one shiny application, and creating a navigation bar and overview of the project.

**7th October**

**Time Spent: 3 hours**

Making a presentation and putting together final documentation for the project, making sure the code is well documented

**Scope for Improvement:**

Improvements to the RShiny project can include enhancing the user interface (UI) for a more intuitive and engaging user experience. Consider incorporating user-friendly features like tooltips, clearer labels, and user instructions to guide visitors through the application. Additionally, providing more detailed descriptions and explanations of the visualizations can help users better understand the insights and data being presented. Clear, concise, and informative descriptions can make the application more informative and accessible to a wider audience.

**Sources and References:**

Written/Online Resources: The resources listed on the course website and course text books were referenced. "I confirm that the above list of sources is complete and we have not talked to anyone else about the solution to the problem"

Chart for Time Log

