The GitGirls Collective

GLOBAL IT SALARY ANALYSIS

Exploring the global IT job salary landscape and analysing the connection between gender-based wage disparities

13 November 2023

SAM KERR, NICOLA PREVOST, TATIANA NGAMBA,
ALICIA MONGE GRASA, SAMANTHA HUGHES-STANLEY

CODE FIRST GIRLS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Aims and Objectives

In 2022 alone, Code First Girls (CFG) gave 44,861 opportunities to women to learn coding, compared to only 6,450 women pursuing undergraduate computer degrees in the UK. CFG is expanding rapidly and has entered foreign markets such as the United States, France, Switzerland, Poland, the Netherlands, and Hungary.

As enthusiastic participants in the CFG Degree program, which is dedicated to facilitating women's transition into the technology sector, our collective interest in technology and IT transcends academic boundaries, influencing our personal and professional pursuits. Driven by a common ambition to secure roles in the tech field, we are eager to explore global opportunities that offer the best working conditions. Our drive stems from an acute awareness of persistent gender-related discrepancies, especially in terms of salary, which we aim to incorporate into our analysis.

Our project aims are:

1. To discover how salaries for IT jobs differ globally.
2. To explore the relationship between population density and salary values through a simple linear regression model.
3. To provide a comprehensive analysis of the global gender pay gap in IT salaries

Roadmap of the report

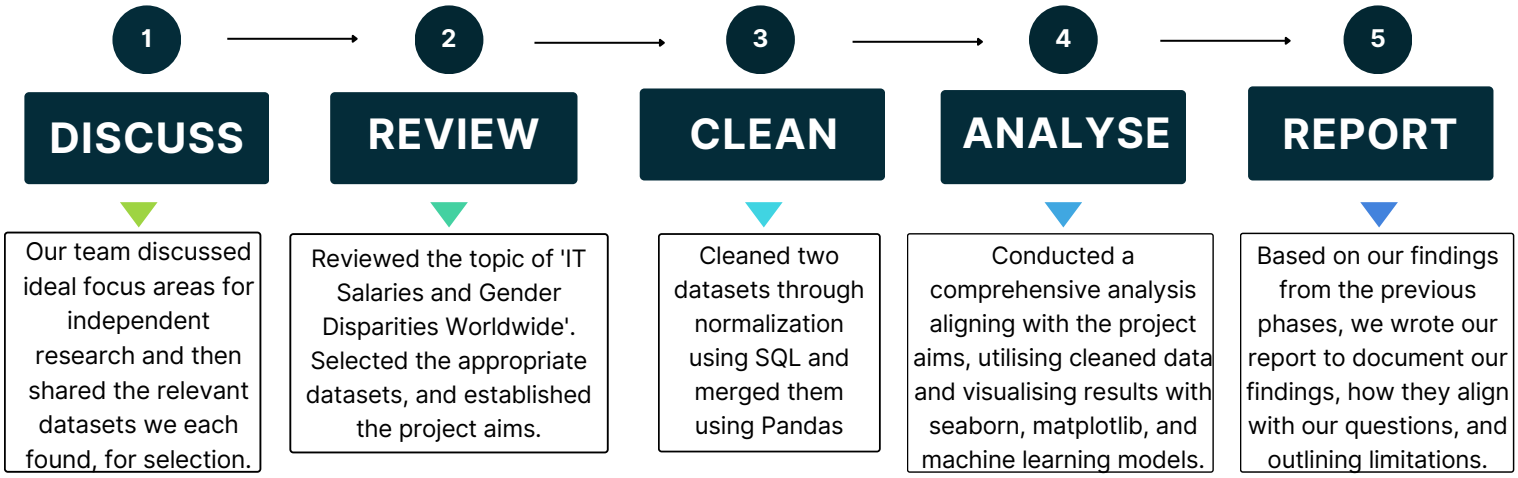


Figure 1: *Project Process Roadmap*

BACKGROUND

In the dynamic landscape of today's technology sector and global job market, understanding the compensation trends for IT professionals is becoming increasingly relevant. This study is particularly pertinent for women contemplating a career shift into technology, as well as for businesses operating in the tech industry. Our analysis offers valuable insights for Business Executives and Managers who are navigating this ever-evolving sector. Our investigation is motivated by the global demand for IT skills, which knows no borders.

Our project was initiated with several compelling inquiries: Would women benefit more from pursuing higher salaried positions in countries with higher average wages but significant gender pay gaps? Conversely, is it more prudent to opt for lower paying roles in countries where income distribution is more equitable? We also explore whether a country's population density or its continental location impacts gender pay parity and average salaries. Furthermore, we examine the influence of gender pay parity on average salaries.

Our research focuses on several key areas:

- The interplay between pay parity, population density, and the continent of a country.
- A comparative analysis of the compensation for IT roles in various countries relative to other professions.
- Identifying the most favourable countries for IT professionals in terms of salary, considering pay parity.
- Using the Support Vector Machine (SVM) model to predict pay parity in different countries.
- Applying the K-Nearest Neighbours (KNN) model to forecast pay parity across nations where data wasn't present in our dataset

STEPS, SPECIFICATIONS & PROJECT DESIGN

Data Gathering:

Our initial step involved gathering data from various sources including Teleport, Geonames, WorldData, and the World Economic Forum. The currency conversion rates and data from the Teleport API were obtained on November 29, 2023. While we attempted to ascertain the exact publication date of the Teleport dataset, our email inquiry remained unanswered. The cost of living data from WorldData is relevant to the year 2023. The publication date for the Geonames data was also unavailable.

Our datasets are formatted as CSV files for ease of manipulation in Jupyter Notebooks. The datasets include:

- **Countries:** This dataset merges details about countries, encompassing gender parity statistics, population figures, ISO alpha-2 codes, and currency codes.
- **Jobs:** Includes information about various occupations from each country in the dataset.
- **Salaries:** This contains the salary information including the conversion rates for the 25th, 50th and 75th percentiles and references both countries and jobs.
- **Cost of living:** This includes information about worldwide cost of living and purchasing power index data per country/region.

Data Cleaning:

We began the data cleaning by constructing a comprehensive spreadsheet using API data, country codes, and gender disparity statistics, with distinct sheets representing SQL tables and sample data. After organising the data for better clarity, an Entity-Relationship (ER) diagram was created to visualise the initial conceptualisation of around 10 tables. A collaborative meeting refined the ER diagram **(Appendix A)**, incorporating primary and foreign keys. Returning to the spreadsheet, we normalised the database into three separate sheets, employing V-LOOKUP formulas for accuracy and efficiency. The SQL-compatible structure was then achieved using an online tool called SQLizer (https://sqlizer.io/), preventing manual formatting.
The **countries** dataset contains 250 rows and 8 columns, the **job** dataset contains 10,297 rows and 3 columns, the **salary** dataset contains 10,297 rows and 9 columns, and the **cost of living** dataset contains contains 99 rows and 5 columns. For the countries dataset, out of the 250 countries considered, gender pay disparity data was available for 136 countries. For the remaining countries, 'NULL' values were populated in the SQL database to replace the absence of this information.
Four specific countries - Antarctica, Bouvet Island, Heard and McDonald Islands, and U.S. Outlying Islands - recorded ‘0’ for their population. Antarctica presented a unique case with 'N/A' for the Currency Code and this was treated as 'NULL' in the SQL database. For the salaries dataset, incorrect values were displayed for an Accountant in Ghana(GH). Please see (**Appendix B**) for the incorrect values screenshot.

Whilst cleaning this data, we realised that we were also missing the converted currency value for the ‘percentile_25_GBP, percentile_50_GBP and percentile_75_GBP’ columns. and this was because the currency code for Belarus, Mauritania and Venezuela were incorrect. These values were missing for ISO Alpha 2 - ‘VE’, ‘MR’ and ‘BY’. Please see **(Appendix C)** for the missing values screenshots.The API code was amended again to retrieve these values for these countries. Whilst the API code was updated, we noticed that the currency code was incorrect hence why these values had not been pulled through. The country code and the salary values were updated in Excel for each relevant country before adding the values in SQL. Please see **(Appendix D)** for the solution screenshots.

Data Analysis and Visualisation:

The team conducted a thorough analysis, exploring how IT roles are compensated in comparison to country medians. We specifically compared salaries in Data Analysis and Software Engineering to the median salary per country, employing Machine Learning algorithms (SVM & KNN) to analyse salary trends along with gender and cost of living factors. This enabled us to investigate potential correlations between pay parity and median salaries across countries. Additionally, geocoding and GeoPandas were utilised for geographical visualisation of pay parity by country. Our analysis incorporates a variety of visualizations, including heat-maps, histograms, box plots and bar charts, to gain comprehensive insights from the data.

Non-technical Requirements:

- **Project aims and objectives -** To set out our project objectives for data analysis.
- **Our target audiences -** Identifying our target audience and how our analysis would be useful to them
- **Project management -** To ensure we are closely answering the project questions.
- **Communication -** To communicate effectively throughout the project.
- **Timeline** - Defining a realistic timeline
- **Team collaboration** - Defining roles and clear expectations for the project.

Technical Requirements:

- **Data sources** - Teleport, Geonames country codes, World Economic Forum gender pay parity data
- **Data Collection** - Python for API and Data Processing and CSV files.
- **Data Normalization & Data Storage** - SQL Database to normalize the data, Excel spreadsheet to understand the data structure and ER Diagram to visualise the data relationships
- **Data Cleaning** - Utilising techniques learned in class to clean the data using Jupyter Notebooks.
- **Data Analysis Tools** - Pandas, NumPy, and other Python libraries were used for data manipulation and analysis using Jupyter Notebooks
- **Data Manipulation** - Utilising Python’s Pandas library.
- **Data Visualisation -** Utilised visualisation libraries such as Matplotlib, Seaborn for effective data representation.
- **Statistical Analysis** - Using statistical methods to gain insights into the dataset (mean, median, correlation, regression and Machine Learning.
- **Project Documentation -** Documenting our project from beginning to end.
- **Report Analysis** - Displayed our visualisations and how our questions have been answered.

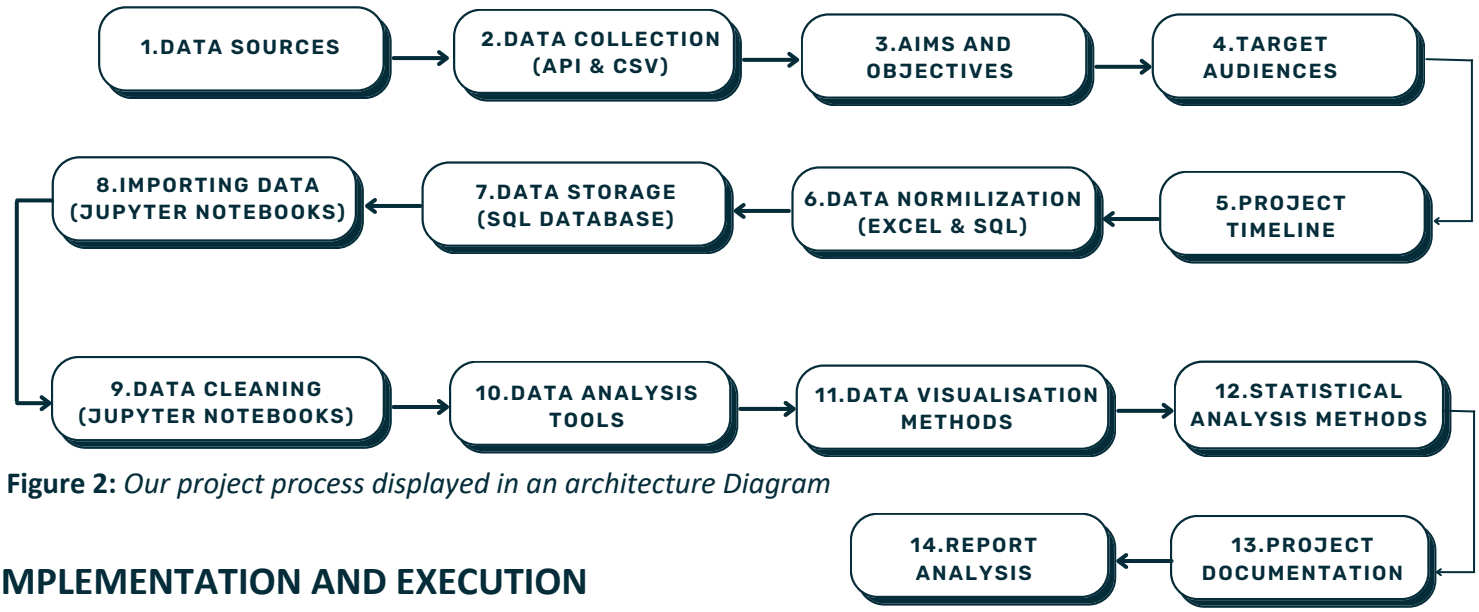
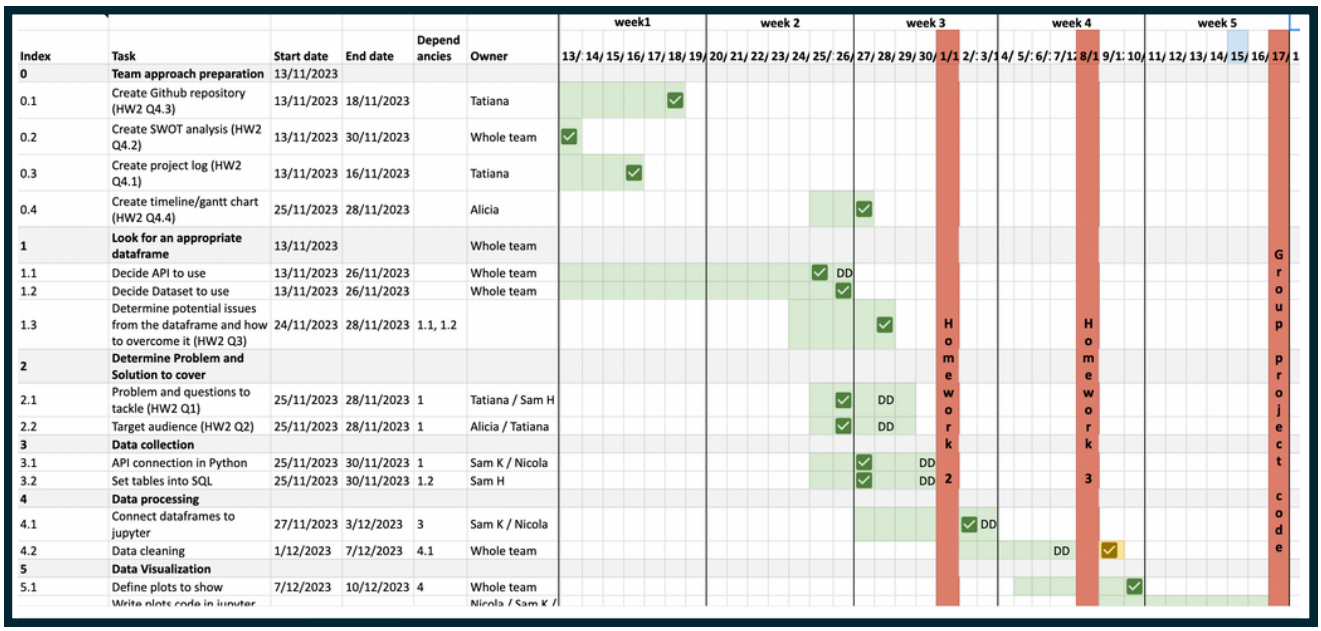


Figure 2: *Our project process displayed in an architecture Diagram*

IMPLEMENTATION AND EXECUTION

As this project had a tight time scale, it was important for us to understand every member's strengths and weaknesses. We conducted a SWOT analysis which is shown below **(Appendix E).**
We took an agile approach for this project, meeting regularly to discuss our progress as a team and delegate tasks. We were able to collaborate flexibly, directing and making changes to the project as we progressed, adapting to each other's feedback. We utilised a Gantt Chart to organise the project timeline and allocate tasks to each team member.

Figure 3: *Gantt Chart of the group project*



Moreover, below are the roles that each member took on for this project. These roles are expanded in detail in the project activity log. The roles were assigned based on each member's strengths and knowledge.

Table 1 : *Team Members roles*

MEMBER NAME	ROLE
Tatiana Ngamba	Data Architect, main Data cleaner and Project Manager
Sam Kerr	Lead Analyst, Data Engineer and Project Coordinator
Nicola Prevost	Lead Data Scientist, Data cleaner, Visualisation and Machine Learning Specialist
Alicia Monge Grasa	Data Analyst and Project Leader
Samantha Hughes-Stanley	Data Architect, Documentation Specialist and Data Integrator

Tools and Libraries:

To successfully execute our project, we utilised a variety of libraries and tools. Some of these were introduced to us during the CFG degree program, while others were identified through online research. Employing these resources enabled us to enhance and elevate our project significantly. On the next page is a table which outlines the libraries and tools we used, detailing the methods of their application and their specific purposes in our project.

Table 2 : *Tools and libraries used by the team*

LIBRARY / TOOL	METHOD	USED FOR
Jupyter Notebooks	To run Python libraries and code and SQL code.	Data Analysis, data visualisation & machine learning
Pandas	pd.read_csv , .head(),.describe(), info, isnull().sum(), .iloc()	Data cleaning , Data Analysis
Matplotlib	pyplot	Data Visualisation
Numpy	median, mean, max, min	Data Analysis
Seaborn	heatmap	Data Visualisation
sqlalchemy	To import the country code table and merge with API DF	Data Normalisation , db connector, importing data
API	To retrieve data from the data source	Importing data
Geo Pandas	Geomap	Data Visualisation
Excel Spreadsheet	To visualise sample data and structure the datasets	Data Normalisation
ER Diagram	To structure our SQL database	Data Normalisation
Warnings	To ignore the warning flags	Data visualisation warnings
CSV	read_csv	Data Analysis
Google Drive	Sharing links to live documents, giving editor permissions to all group members, for collaboration	Project Management, To access project requirements, project activity, homework two and project documentation
GitHub	Sharing documents efficiently, ensuring the most recent changes were available for the group to download and edit	To update the latest version of our documents/code.
Zoom & Google Meet	Meeting links shared in Slack	Team meeting, Project management, collaboration and communication.
Slack	Updating the group about progress and relevant information on our group channel	Communication and sharing ideas and progress
Gannt Chart & Github Project	Created in Excel and GitHub, incorporating our project's requirements	Project Management, Project timeline

DATA COLLECTION OVERVIEW

Our project necessitated collecting salary data across various countries, especially focusing on IT roles. We required a straightforward metric for gender pay disparity, and, if time permitted, additional country-level statistics like population or land area. Our progress and research methodology were meticulously documented in the **'Group 7 Project Requirements Breakdown'** spreadsheet, which included a dedicated section for dataset links.

Our project was primarily structured around the data accessible from Teleport.org's developer API, providing detailed country information (like population, currency code, capital, etc.) for 252 countries. Out of these, the salary data (covering the 25th, 50th, 75th percentiles) for 52 job roles was available for 198 countries.

We referred to geonames.org for consistent iso_alpha2 codes, as recommended by Teleport.org, ensuring uniformity in our data. This site also served as our go-to source for broader country-level statistics.For additional data, we selected sources based on the credibility of the providing organization, such as WorldData and the World Economic Forum.

Table 3 : *Data sources, the retrieval method and the filenames generated.*

Data Source	Retrieval Method	Filename
1.Teleport.org	**Python code API calls x 3**	**output_gbp_salaries_{timestamp}.csvFrozen version: output_gbp_salaries_23-11-29_10-55.csv**
https://teleport.org/	simple_get_list_countries() - retrieved a simple list of countries for which Teleport offers data (country name and hyperlink to API endpoint)	**list_countries.json**
https://developers.teleport.org/api/resources/Country/	get_overviews_countries() - made a series of API calls to retrieve country-level overview data (most importantly, currency code and iso_alpha2 code)	**overviews_all_countries.json**
https://developers.teleport.org/api/resources/CountrySalaries/	api_to_dataframe()retrieves country-level salary data, utilising JSON data retrieved from the previous two API requests.	**output_inc_codes.csv**
2.Exchangerate-api.com	**Python API call within**	
https://www.exchangerate-api.com/ | https://open.er-api.com/v6/latest/{currency}	get_conversion_rates_insert_df()utilises the currency codes of the countries we are researching to make an API request for the currency exchange rate against GBP as a base.	**currency_rates_{timestamp}.json**
3.Geonames.org. https://www.geonames.org/countries/	Replicated the table from the source's website by creating an equivalent SQL table with suitable headings, data types, and keys, and utilised SQLizer, an online tool, to format the data into SQL-compatible structure, streamlining the process and eliminating the need for manual formatting. https://sqlizer.io/	**country_codes.sql**
4. World Economic Forum https://www.weforum.org/publications/global-gender-gap-report-2022/in-full/1-benchmarking-gender-gaps-2022/	Extracted the country and gender pay parity data from the website, then reformatted it into a CSV format. This streamlined the integration of the data into our analysis, ensuring efficiency and accuracy.	**Gender Pay Gap.csv**
5.WorldData https://www.worlddata.info/cost-of-living.php	Table directly copied into csv from website. Used excel VLOOKUP function to add relevant iso_alpha2 codes from output_gbp_salaries{}.csv	**overviews_all_countries.json**

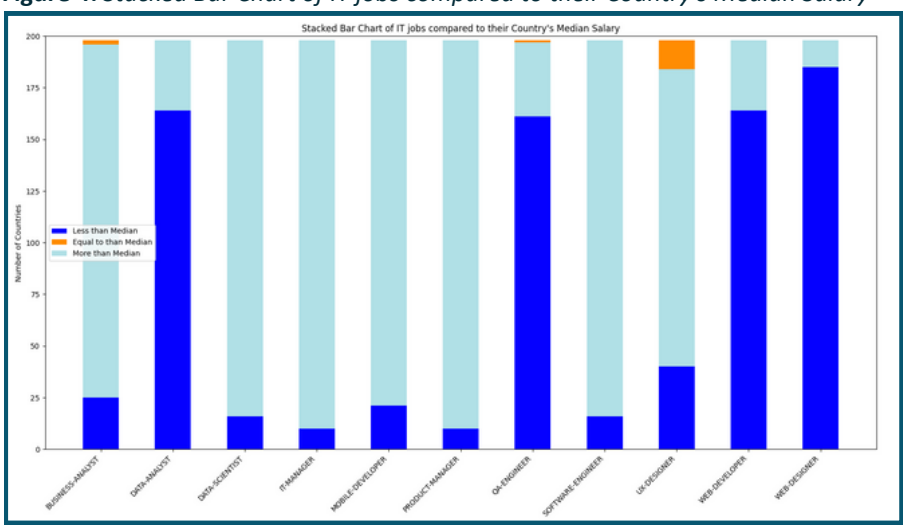
RESULT REPORTING - KEY FINDINGS AND INSIGHTS

QUESTION 1: To discover how salaries for IT jobs differ globally.

Section 6.1 - *How are IT roles paid in comparison to the country medians?*

To examine this question, we used a stacked bar chart to compare IT jobs to their country's median salary, we found that Data Analysts, QA Engineers, Web Developers and Web Designers stand out as occupations where the salary is very likely to be compensated below the country's median salary. Business Analysts, Data Scientists, IT Managers, Mobile Developers, Product Managers, Software Engineers and UX Designers stand out as occupations where the salary is very likely to be compensated above the country's median salary. Another interesting observation is the contrast between the extremes, and that very few countries have IT salaries that are exactly at the median, indicating a clear divide between higher and lower-paying roles.

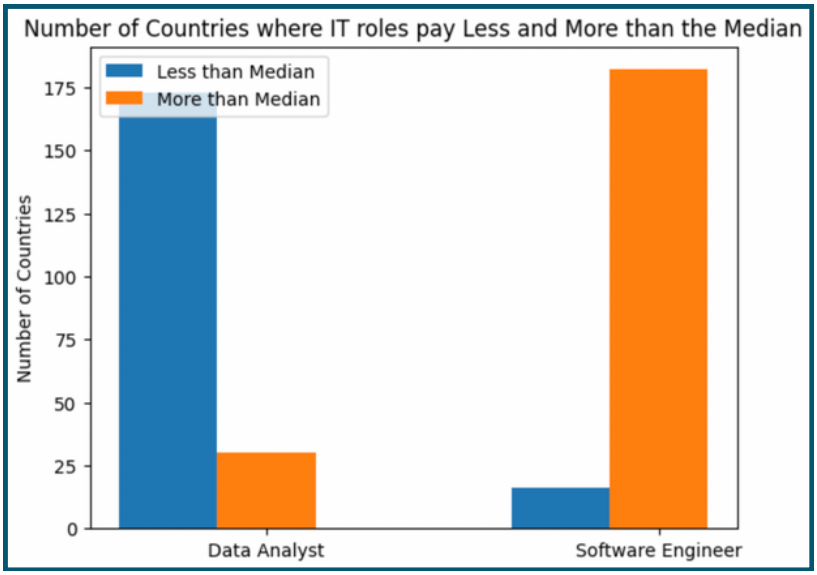
Figure 4: *Stacked Bar Chart of IT jobs compared to their Country's Median Salary*



Section 6.2 - How do the salaries of Data Analysis compare to those of Software Engineers?

To answer this question, we used a bar chart to visualise the number of countries where Data Analysts and Software Engineers earned either more or less than the median salary per country, we can clearly observe that Software Engineering is a higher-paying role compared to Data Analysis in the context of our data. There is a stark contrast between the two professions in terms of compensation, with Software Engineers being better paid in the majority of countries. If you want a greater choice of where to live in the world whilst being well paid, choose to study Software Engineering rather than Data Analysis.

Figure 5: *Number of Countries where IT roles pay Less and More than the Median*



QUESTION 2: Explore the relationship between population density and salary values through a simple linear regression model

Section 6.4: Is Pay Parity Correlated with Continent Population Density?

Firstly we created two scatter plots comparing the population density, pay parity and continent of each country. From this we could not see a clear correlation between these variables, however through further analysis using machine learning, we did discover some significant patterns. In the initial plots we found that:

Asian countries show a wide range of pay parity and population densities. Some Asian countries have high population densities but span across the range of pay parity.European countries tend to cluster in the mid to high range of pay parity. Population density varies, with a few outliers having very high population density. African countries generally have lower pay parity, with population densities spread across the range.South American data points are mostly in the lower half of population density with pay parity values mostly in the middle range.There are fewer data points to compare for Oceania, but they tend to have lower population densities and vary in pay parity.North American countries appear to have a spread in population density but generally higher pay parity.There does not appear to be a clear correlation between population density and pay parity across the continents, however through other analysis methods we did discover firmer patterns. Developed regions (like North America and Europe) tend to have higher pay parity regardless of population density.

We explored this topic further using an SVM machine learning model, which identified an accuracy score of 0.68, which is fairly strong. This indicates that there is a significant correlation between pay parity and continent population density.

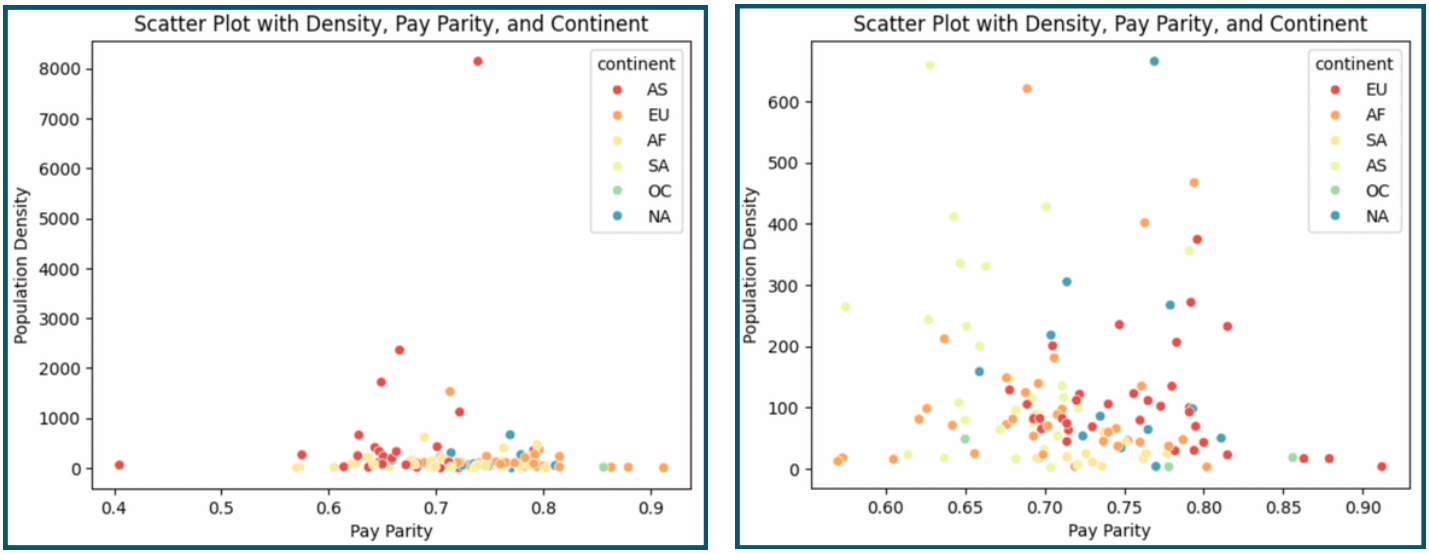
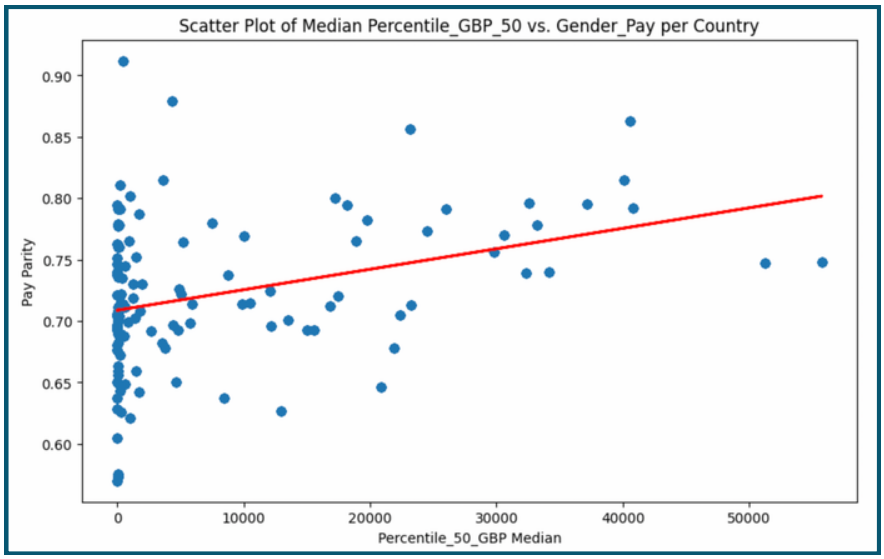


Figure 6: *Scatter Plot with Density, Pay Parity, and Continent*

Section 6.5: Is there is a relation between pay parity and the economic success of a country?

For this question we created a scatter plot with a line of best fit, mapping out the relationship between these factors. From this we could see there is a slight positive correlation between the pay parity and the median salaries per country. There is a clustering of countries at the lower end of the income scale, indicating that a large number of countries with lower median incomes have varying gender pay ratios. From this we can hypothesise that while there is a relationship between the median income of a country and its gender pay gap, it is likely influenced by a variety of other factors.

Figure 7: *Scatter Plot of Median Percentile GBP 50 vs. Gender Pay per Country*

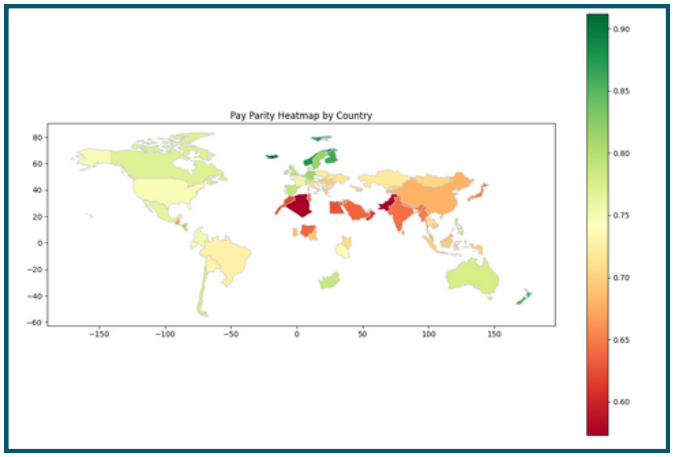


QUESTION 3: Provide a comprehensive analysis of the global gender pay gap in IT salaries.

Section 6.3: How Does Pay Parity Vary Across the World?

Using a geo-heatmap, we were able to visualise how pay parity varies across the world.
The countries which appeared most red in colour have the largest gender pay gaps, while the countries in green have a smaller gender pay gap.
Most countries fell somewhere in the middle of the scale, as indicated by the prevalence of yellow and orange shades. Using this method of visualisation, we can see at a glance that certain regions' countries have similar pay parity levels, which could be indicative of regional economic patterns, labour laws, cultural norms, or the existence of gender equality initiatives.

Figure 8: *Pay Parity Heatmap by Country*



Section 6.6: Is there a linear relationship between gender pay (dis)parity and cost of living metrics or population?

OLS Regression Results			
Dep. Variable:	gender_pay_parity	**R-squared:**	0.412
Model:	OLS	**Adj. R-squared:**	0.388
Method:	Least Squares	**F-statistic:**	17.30
Date:	Sat, 16 Dec 2023	**Prob (F-statistic):**	1.30e-08
Time:	20:15:36	**Log-Likelihood:**	121.29
No. Observations:	78	**AIC:**	-234.6
Df Residuals:	74	**BIC:**	-225.1
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.8329	0.027	30.799	0.000	0.779	0.887
WD_cost_living_rank	-0.0018	0.000	-5.450	0.000	-0.002	-0.001
population	-4.038e-11	2.71e-11	-1.489	0.141	-9.44e-11	1.37e-11
purchasing_power_index	-0.0004	0.000	-1.324	0.190	-0.001	0.000

Omnibus:	0.982	**Durbin-Watson:**	2.021
Prob(Omnibus):	0.612	**Jarque-Bera (JB):**	0.670
Skew:	-0.225	**Prob(JB):**	0.715
Kurtosis:	3.065	**Cond. No.**	1.06e+09

Table 4,5,6 : *OLS Regression Results*

To answer this question we used OLS Regression Analysis. From this we found an R-squared value of 0.4, showing only a weak correlation between gender pay parity and the variables. However, the p>t values for population and WD cost of living rank were significant, indicating that there was in form of relationship. Upon visualising the data using a scatterplot, a negative correlation was observed for the cost of living rank and a positive correlation was observed for population density.

Section 7.1: Prediction of Pay Parity category with population and continent number as explanatory variables

Table 7: *Metrics measure the performance of the model*

Accuracy score	**0.68**
Mean Squared Error	**0.32**
Mean Absolute Error	**0.32**
Precision	**0.68**
Recall	**0.68**

The model has an accuracy score of 0.68 which suggests that it is a fairly accurate prediction model.
The MSE and MAE are relatively low which indicates that there were few errors made in predicting the pay parity category.
The precision and recall were 0.68 which is indicative of an above average prediction model. Precision measures the correct predictions made by the model, where the maximum is 1.0. Recall measures the relevant data points that were correctly identified by the model (the maximum is also 1.0).
Overall, this SVM model is a good* prediction model.
In summary, this SVM model has a moderate level of accuracy, precision, and recall. The MSE and MAE values indicate some level of error in the predictions, but the model is overall reasonably effective in making true predictions.

CONCLUSION

1. Project Limitations: Our project encountered several challenges. A notable limitation was the non-standard distribution of salary data, necessitating additional analysis. Moreover, the data on pay parity was incomplete for some countries because different data sources often had varying criteria for classifying what constitutes a country. Data from Teleport also showed inaccuracies and inconsistencies, especially regarding salary figures, and there was a marked lack of detailed information on the origins and specifics of this data. We attempted to contact Teleport to iron this out, however we did not receive a response. Teleport claimed to have gathered their data from highly reputable sources, however the data appeared to be flawed in some aspects, and due to time restraints of the project, we needed to continue and provide the most accurate analysis we could, with the resources available.

2. Project Strengths: Despite these challenges, several elements of the project were notably successful. Our implementation of Machine Learning models, particularly the SVM with selected variables, was effective. The API for currency conversion was utilised efficiently, and the visualisations created were highly pertinent to our hypotheses. Our SQL implementation and overall time management were commendable. Furthermore, our adaptability in addressing data limitations was a key strength of the team.

3. Quality Data Sourcing: In terms of data sourcing, we took a conscientious approach, prioritising the quality and relevance of the data. Our initial brainstorming sessions, focusing on social issues such as gender parity and cost of living, guided our selection of data topics. We opted for Teleport as our main data source due to its comprehensive data offering on various quality-of-life metrics, and its claimed reliability and reputability.

4. Going Beyond Basic Analysis: Our project aimed to extend beyond basic analysis. We endeavoured to broaden Teleport's city-level data to a more comprehensive country level comparison. This approach allowed us to incorporate gender-pay parity metrics, thereby enriching our analysis and offering deeper insights into the data.

5. Due Diligence: We conducted thorough due diligence in investigating Teleport's data. This involved a detailed examination of their API documentation and other data sources to assess data quality and methodology. We also contextualised the salary data by including additional metrics such as cost of living and purchasing power, adding depth to our analysis.

6. Focused Research Questions: Our research questions were meticulously crafted to be both incisive and attainable. We explored the impact of gender pay gaps on career choices in different countries, with a specific interest in IT career salaries. This focus allowed for a more targeted and detailed analysis, providing valuable insights for our audience.

7. Building an API-Driven Dataset: A pivotal aspect of our project was constructing a dataset driven by API requests. We developed a Python programme to retrieve and build a comprehensive dataset, reflecting the most current data available on Teleport's platform, including real-time currency conversion rates.

8. Future Directions: Looking forward, we would aim to expand our analysis to include a wider range of job roles, to provide more general and useful findings for a wider audience. We would plan to refine our Machine Learning models for more nuanced insights and seek additional data sources to fill existing gaps and correct inconsistencies, thereby enhancing the accuracy and quality of our analysis.

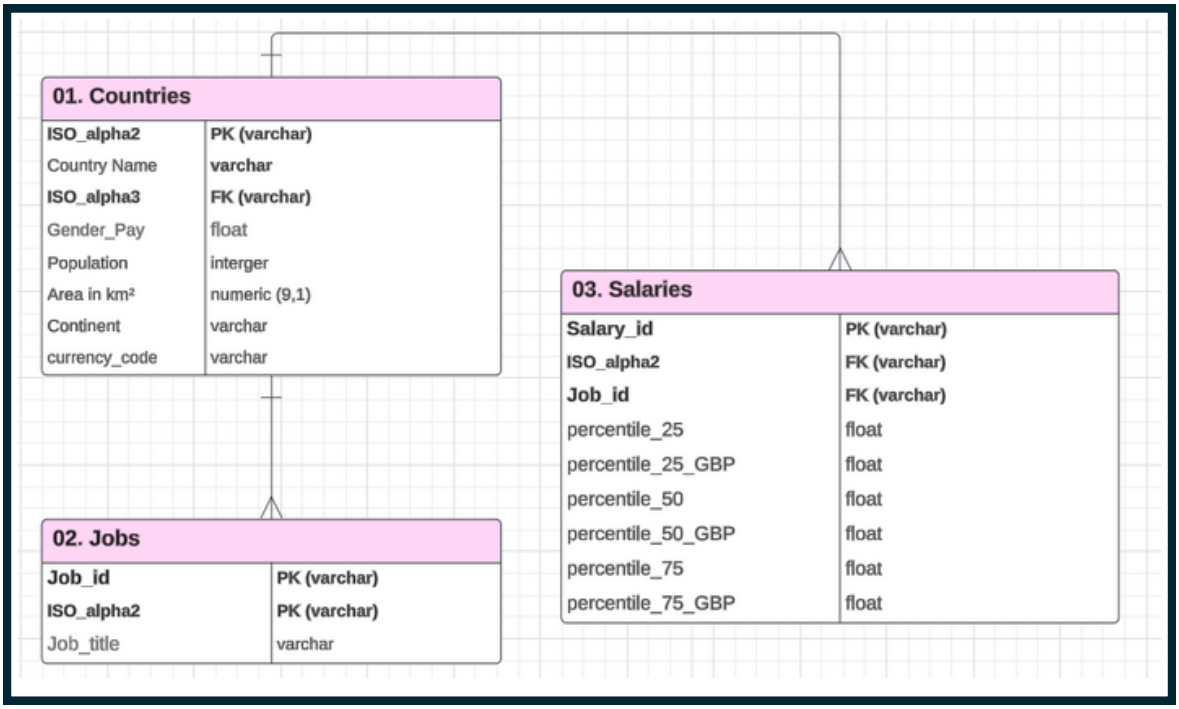
REFERENCES

- **Code First Girls Website** (2023) *CFG*. Available at: https://codefirstgirls.com/blog/cfg-100k-women-taught-to-code announcement/#:~:text=The%20gender%20gap%20is%20closing,every%20115%20roles%20by%202025. [Accessed: 16th December 2023].
- **Teleport (2023)** ***Teleport for Developers***. Available at: https://developers.teleport.org/ [Accessed: 25th November 2023].

APPENDICES

Appendix A:

An Entity-Relationship (ER) diagram was created to visually represent the conceptualisation of SQL tables for data normalization, incorporating primary and foreign keys.



Appendix B:

Incorrect value in salaries dataset for an accountant in Ghana(GH).

job_title	percentile_25	percentile_25_GBP	percentile_50	percentile_50_GBP	percentile_75	percentile_75_GBP
Accountant	1	0.065696677	2	0.131393354	4	0.262786709

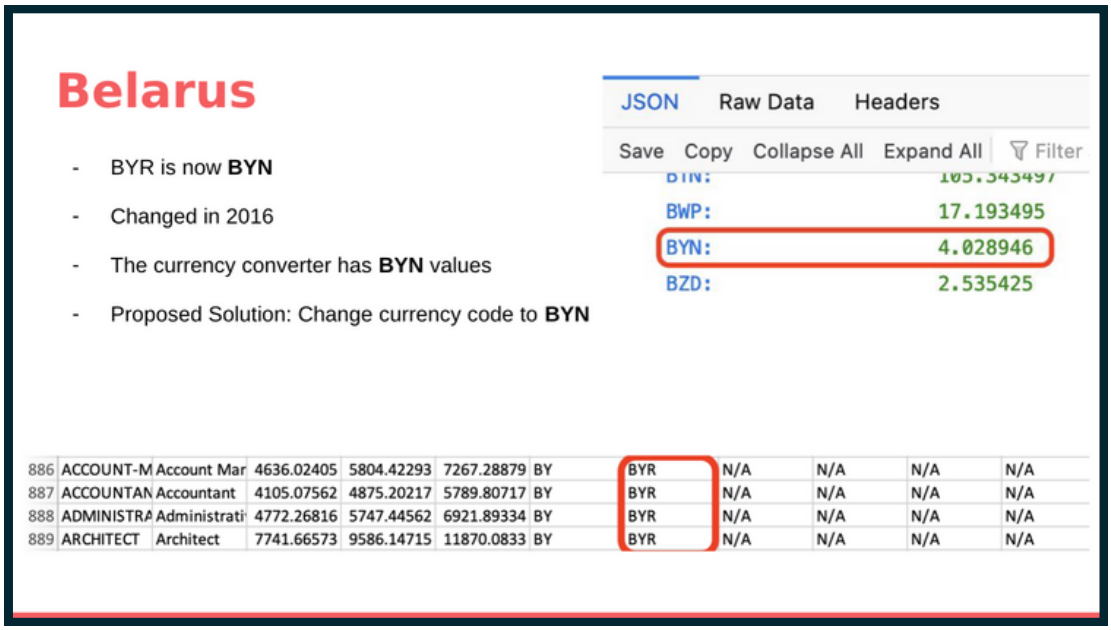
Appendix C:

The missing converted values for the salary table

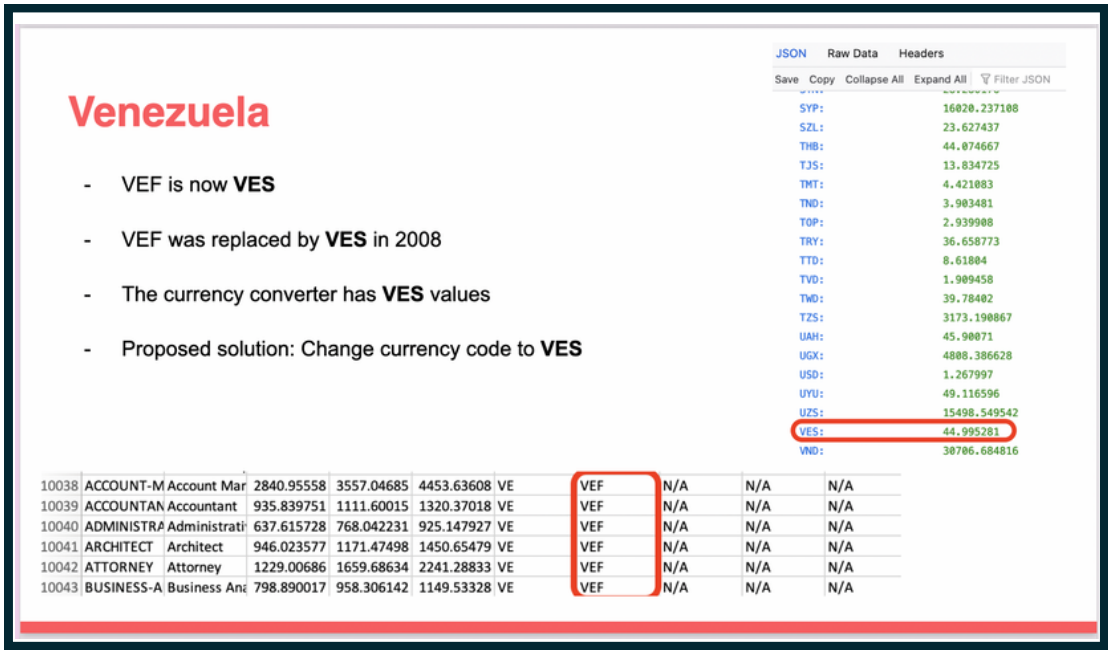
iso_alpha	job_id	job_title	percentile_25	percentile_25_GBP	percentile_50	percentile_50_GBP	percentile_75	percentile_75_GBP
MR	OPERATIONS-MANAGER	Operations Manager	5561.0939	N/A	7087.466933	N/A	9032.788949	N/A
MR	PHARMACIST	Pharmacist	10734.86845	N/A	11959.03264	N/A	13322.79592	N/A
MR	PHYSICIAN	Physician	13513.49913	N/A	18411.92163	N/A	25085.94219	N/A
MR	POSTDOCTORAL-RESEARCHER	Postdoctoral Researcher	4439.875773	N/A	5021.529175	N/A	5679.38306	N/A
MR	PRODUCT-MANAGER	Product Manager	7589.908271	N/A	9327.270774	N/A	11462.32299	N/A
MR	PROJECT-MANAGER	Project Manager	6799.676409	N/A	8405.079134	N/A	10389.51724	N/A
MR	QA-ENGINEER	QA Engineer	5358.602502	N/A	6605.866512	N/A	8143.442692	N/A
MR	RECEPTIONIST	Receptionist	2773.036102	N/A	3358.607099	N/A	4067.830792	N/A
MR	RESEARCH-SCIENTIST	Research Scientist	5803.495975	N/A	7437.984015	N/A	9532.806853	N/A
MR	SALES-MANAGER	Sales Manager	5795.431828	N/A	7659.889485	N/A	10124.1648	N/A
MR	SOFTWARE-ENGINEER	Software Engineer	6820.734274	N/A	8313.294365	N/A	10132.46675	N/A
MR	SYSTEMS-ADMINISTRATOR	Systems Administrator	5481.805118	N/A	6597.542569	N/A	7940.371285	N/A
MR	TEACHER	Teacher	3919.715024	N/A	4848.070648	N/A	5996.300461	N/A
MR	UX-DESIGNER	UX Designer	6037.074562	N/A	7402.430802	N/A	9076.5786	N/A
MR	WAITER	Waiter	1578.519737	N/A	2416.290984	N/A	3698.694403	N/A
MR	WEB-DESIGNER	Web Designer	4392.743435	N/A	5375.049729	N/A	6577.019583	N/A
MR	WEB-DEVELOPER	Web Developer	5534.773284	N/A	6858.173174	N/A	8498.00649	N/A
VE	ACCOUNT-MANAGER	Account Manager	2840.95558	N/A	3557.046847	N/A	4453.636079	N/A
VE	ACCOUNTANT	Accountant	935.8397512	N/A	1111.600153	N/A	1320.37018	N/A
VE	ADMINISTRATIVE-ASSISTANT	Administrative Assistant	637.6157278	N/A	768.0422311	N/A	925.1479269	N/A
VE	ARCHITECT	Architect	946.0235774	N/A	1171.474983	N/A	1450.654792	N/A
VE	ATTORNEY	Attorney	1229.006862	N/A	1659.686337	N/A	2241.288329	N/A
VE	BUSINESS-ANALYST	Business Analyst	798.8900167	N/A	958.3061423	N/A	1149.533281	N/A
VE	BUSINESS-DEVELOPMENT	Business Development	708.8320277	N/A	939.1294318	N/A	1244.249773	N/A
VE	C-LEVEL-EXECUTIVE	C Level Executive	964.9775393	N/A	1447.141505	N/A	2170.225161	N/A
VE	CASHIER	Cashier	154.2983561	N/A	216.1889702	N/A	302.9045288	N/A
VE	CHEF	Chef	636.1506392	N/A	781.6459862	N/A	960.4178791	N/A
VE	CHEMICAL-ENGINEER	Chemical Engineer	658.3301407	N/A	797.7403459	N/A	966.6725252	N/A
VE	CIVIL-ENGINEER	Civil Engineer	800.246249	N/A	954.2610273	N/A	1137.917372	N/A
VE	CONTENT-MARKETING	Content Marketing	1027.593912	N/A	1286.656252	N/A	1611.0297	N/A
VE	COPYWRITER	Copywriter	719.3098033	N/A	884.559998	N/A	1087.773845	N/A
VE	CUSTOMER-SUPPORT	Customer Support	455.662311	N/A	586.0382676	N/A	753.7179241	N/A

Appendix D: Missing currency values & solution update the currency code

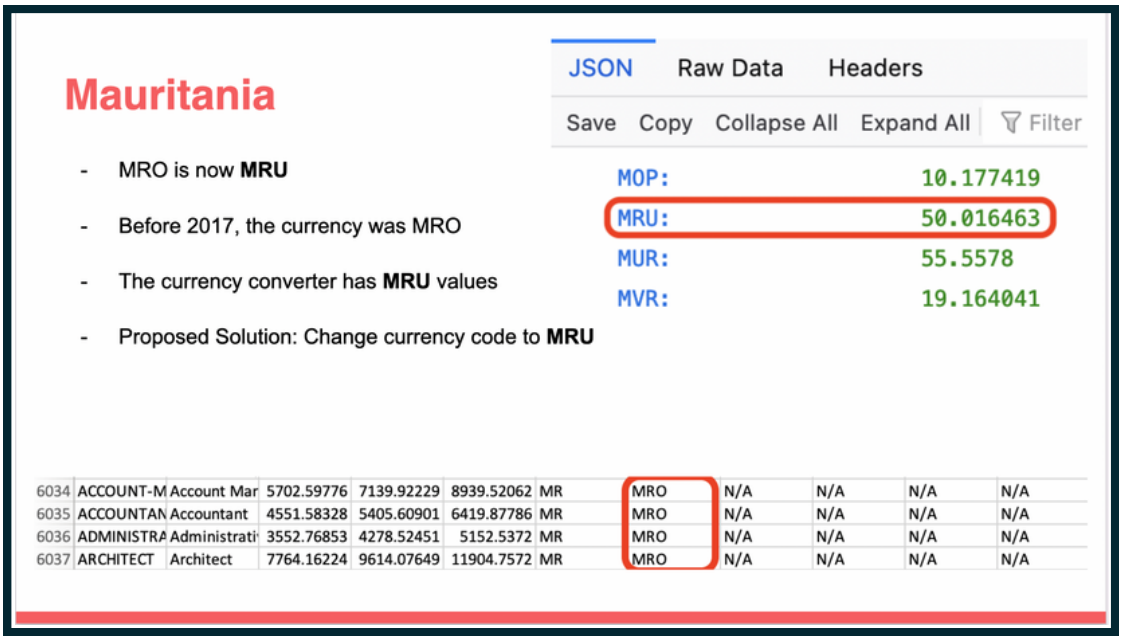
Explanation for one of the missing value for BY (BYR).



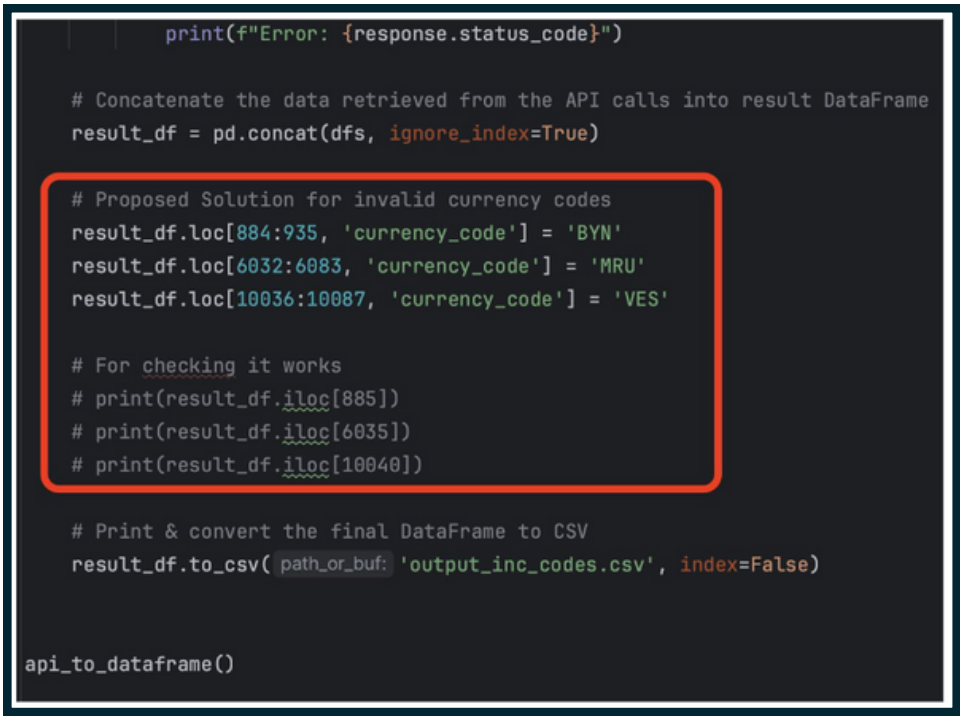
Explanation for one of the missing value for VE (VEF).



Explanation for one of the missing value for MR (MRO).



Solution to fix the missing currency value by updating the currency code to BYN,MRU and VES



```
            print(f"Error: {response.status_code}")

    # Concatenate the data retrieved from the API calls into result DataFrame
    result_df = pd.concat(dfs, ignore_index=True)

    # Proposed Solution for invalid currency codes
    result_df.loc[884:935, 'currency_code'] = 'BYN'
    result_df.loc[6032:6083, 'currency_code'] = 'MRU'
    result_df.loc[10036:10087, 'currency_code'] = 'VES'

    # For checking it works
    # print(result_df.iloc[885])
    # print(result_df.iloc[6035])
    # print(result_df.iloc[10040])

    # Print & convert the final DataFrame to CSV
    result_df.to_csv( path_or_buf: 'output_inc_codes.csv', index=False)


api_to_dataframe()
```

Appendix E:

The team conducted a **SWOT Analysis** at the beginning and end of the project to evaluate strengths, weaknesses, opportunities, and threats.

Name	Strenghts	Weaknesses	Opportunities	Threats
Sam Kerr	**Documentation,Organization, Project Management, Python**	ML, AI,Appreciate if work tasks allocated	CyberSec industry / concepts familiarity	Away Thur 30th Nov - Mon 4th Dec
Samantha H	**SQL, Documentation, Analysis, Communication**	Less experience in Python and the relevant libraries. Prefers collaborative work where possible, over all independent work	Specialist knowledge on careers guidance, as is an Employment Advisor	
Tatiana	**SQL, Project Management, Attention to detail, time management, able to process large data sets, documentation of processes. Pandas, jupyter notebooks**	Data Visualisation, taking on too much work if work is not allocated , ML	Fashion, Education, Recruitment	
Nicola	**ML, jupyter, pandas, statistics**	Writing clearly and coherently.	Chemical and pharmaceutical interest, Criminology	
Alicia	**Excel (from analysis to visualizations), SQL, Pandas, Project Management, Attention to detail, critical and analytical thinking**	ML, API calls, appreciate if work tasks allocated	Knowledge on the field od science, specifically in Food science	
Sarra	**SQL**	Struggle with group callsAppreciate if work tasks allocated		Drop out of the course