PROJECT PORTFOLIO

2022

SURVEY OF DATA PROFESSIONALS
BY OLAKUNLE ADEGEYE

(Microsoft SQL, PowerBI, Tableau, Excel, DAX, Figma)

A data analysis project created to showcase my skills in

Microsoft SQL, PowerBI, Tableau, Power Query,

DAX, Excel, data reporting, analytical skills

Communication and presentation skills

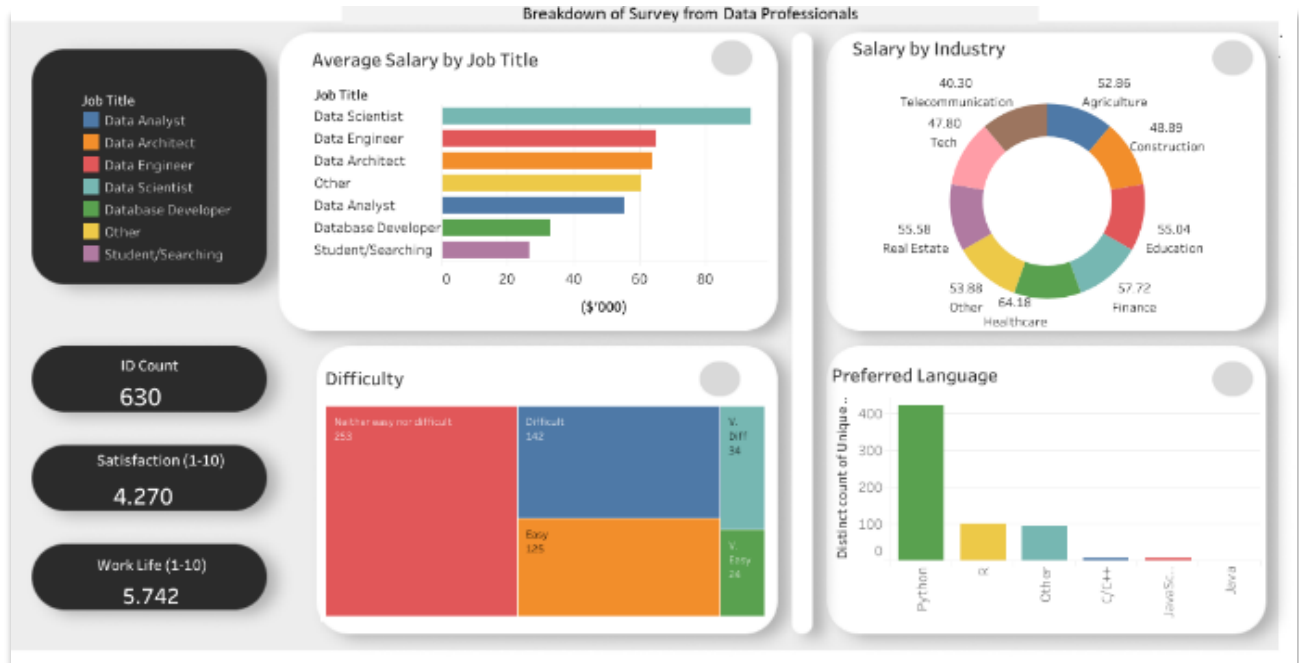# TABLE OF CONTENT

PRELIMINARY DASHBOARD VIEW

FIG 1: TABLEAU DASHBOARD (Click **HERE** for interactive version)
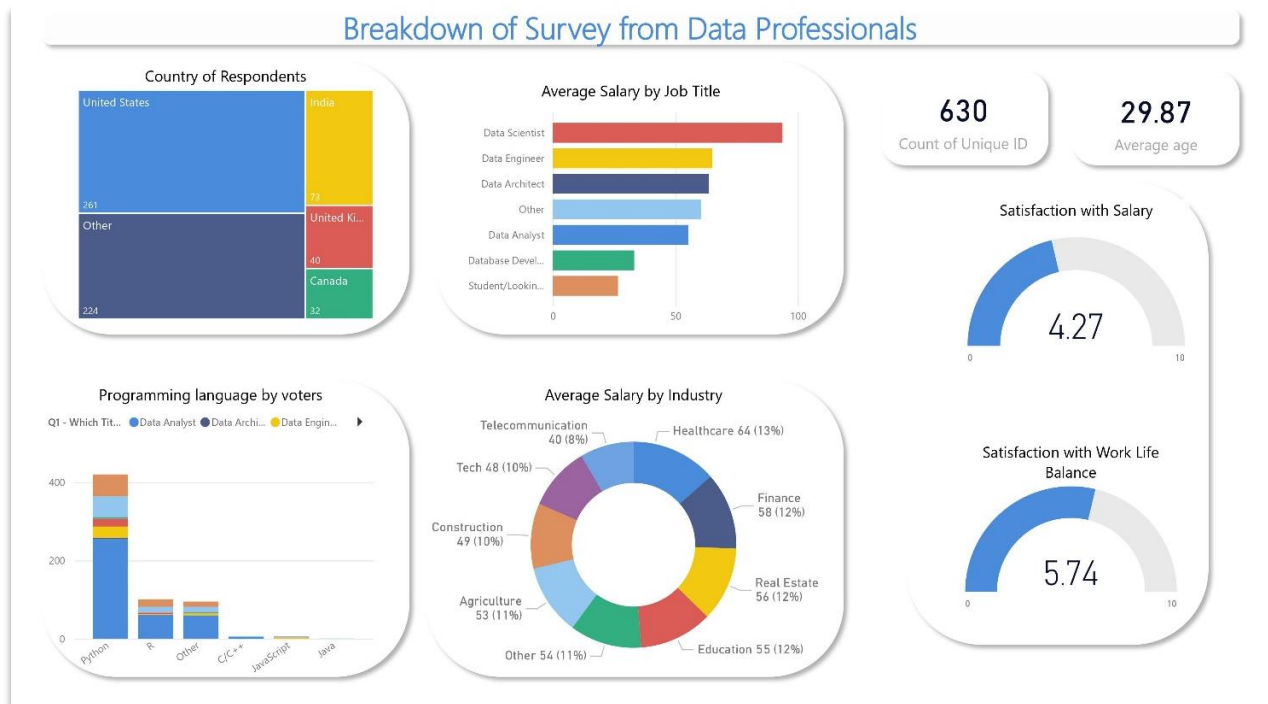

FIG 2: POWERBI DASHBOARD (Click **HERE** for interactive version)

TECHNOLOGY STACK USED:

- Microsoft SQL Server (for data cleaning, transformation and summary)
- PowerBI (for dashboard presentation)
- Excel (For Data cleaning, transformation and summary)
- Tableau (for dashboard presentation)
- Figma (for dashboard design)
- PowerPoint (for dashboard design)

SKILLS HIGHLIGHTED:

- Data Cleaning and transformation using SQL, DAX functions, Power Query and Excel
- Data summarization and manipulation using SQL and DAX
- Data visualization using PowerBI, Tableau and Figma
- Communication skills
- Analytical skills

RESOURCES AND RESULTS

- MSSQL Script for cleaning and transformation HERE
- Raw dataset HERE
- Power Query cleaned dataset HERE
- Tableau Dashboard HERE
- PowerBI Dashboard HERE

EXECUTIVE SUMMARY

A descriptive research was carried out amongst data professionals to understand various aspects about the profession such as the average salary, most preferred programming language, ease of transitioning to the field, among other questions. The major objective of the analysis was to better inform aspiring data professionals about the profession with respect to the effort and reward.

Data Scientist, Data Engineers, Data Analysts, Data Architects, Database Developers comprised majority of the respondents and purposive sampling was used to select a number of social media platforms for the distribution of the survey form. Data was collected using an online survey form and the responses were downloaded as a spreadsheet file.

Data cleaning, transformation and summarization was done using various tools such as Microsoft SQL Server, DAX functions and Power Query, and Microsoft Excel.

The results of the data were presented in a dashboard view using Tableau and PowerBI. This was done to showcase my proficiency with both BI solutions. The links to the interactive PowerBI dashboard can be accessed by clicking HERE, while the link to the Tableau dashboard can be accessed HERE. Static summaries of the dashboards were also provided towards the end of this report.

The result of the analysis revealed that Data Scientist made the most money compared to other data professionals with an average salary of $93,800. This resonates with other industry analysis carried out using job boards such as Glassdoor and Indeed. The Finance sector was discovered to be the most lucrative sector according to the analysis. Python was chosen as the most preferred programming language with 67% of the respondents preferring it over other programming languages. 72% of responses as regards transitioning into the field ranged between 'very easy' and 'neither easy nor difficult'. Nevertheless, the satisfaction with salary in the field came in at 4.27 out of a total score of 10.

The research had certain limitations: majority of the responses were from Data Analyst. SQL was excluded from the list of languages, and the report did not cover the entire breadth of the analysis possible.

It is advised that anyone interested in the information or considering transitioning into a profession in data should interact with the dashboard to fully understand the analysis.

PROBLEM STATEMENT.

Data related professions have been experiencing tremendous growth in line with the increase in data being created. Data has been growing at such an exponential rate that almost every year, we hear claims that 90% of the world's data was generated in the past 2 years. Companies are becoming more data conscious and using it to create competitive advantage. However, a large percentage of the data created is usually unstructured, hence the need for data professionals.

The objective of the analysis is to describe the status quo in the industry with respect to the favorite programming languages, salary ranges, job satisfaction among other factors. The result of the analysis will be used to provide some recommendations and also promote the growth and easy transition into the field.

RESEARCH DESIGN.

The research is both descriptive and prescriptive. Descriptive because it attempts to understand the current situation. It's also prescriptive because the analysis ends with a recommendation for aspiring and incumbent data professionals.

POPULATION AND SAMPLE

The population of the study were data professionals. The sample comprised Data Scientist, Data Engineers, Data Analysts, Data Architects, Database Developers, other professionals in the field of data classified as 'other'; and students alongside people looking for jobs in the field classified in the same category.

SAMPLING TECHNIQUE

The research employed both purposive and simple random technique. Purposive sampling technique was used in the selection of the platform for distributing the questionnaires. Twitter, Instagram and a few other social media platforms were used for distribution. Simple random was used as regards the openness to all data professionals who were willing to participate within the selected platforms.

DATA COLLECTION METHOD

Data was collected using an online survey form. After the forms were the filled, the data was downloaded as a spreadsheet file. Click HERE to download the dataset

DATA CLEANING AND TRANSFORMATION

The data cleaning and transformation was done using:

- Microsoft SQL Server
- Power Query and DAX functions in Microsoft PowerBI.
- Microsoft Excel

SQL was used to perform some aggregate functions and summarize the data. It was also used to clean and transform the data that was used to create a dashboard with Tableau.

Power Query was used to clean and transform the data that was used to create a dashboard with PowerBI.

As part of the cleaning process, the irrelevant and empty columns were removed from the dataset.

Since the research intended to focus on a limited set of professionals, the title column was cleaned to include only Data Analyst, Data Scientist, Data Engineer, Database Developer, and other data professionals.

The split column function was used to separate the irrelevant data by splitting it at the left parenthesis, Favorite programming language column was split at the colon leaving the column with Python, R, C/C++, Java, Javascript and 'other'

The salary range was presented as 00k – 000k which would have made it difficult for analysis. The column was split into the upper and lower boundary of the salary, and the datatype was transformed from string to integer, after which a DAX formula was used to get the mean value of the upper and lower boundaries of the salaries.

Similar actions were carried out on other columns that contained irrelevant information, before the data was loaded into the visualization pane
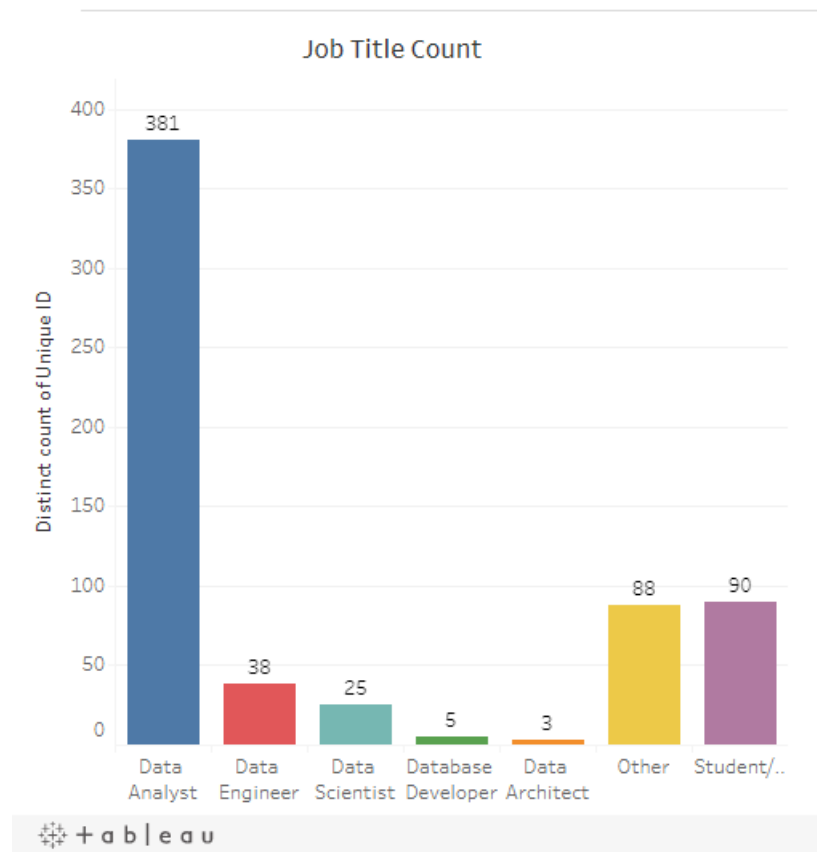
RESULTS

Fig 3: JOB TITLE COUNT (Designed with Tableau)

There were 630 respondents who participated in the survey. 381 were Data Analyst, 38 were Data Engineers, 25 where Data Scientists, 5 were Data base developers, 3 were Data Architects, 88 were unclassified, and 90 were either students or people searching for jobs in the field.
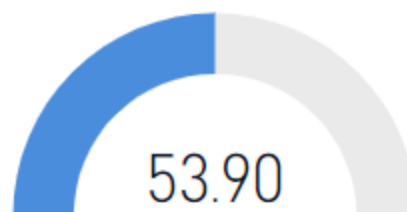

FIG 4: AVERAGE SALARY (Designed with PowerBI)

According to the report, the average salary in the industry was $53,900. The minimum salary was $40,000, while the maximum value was $225,000.
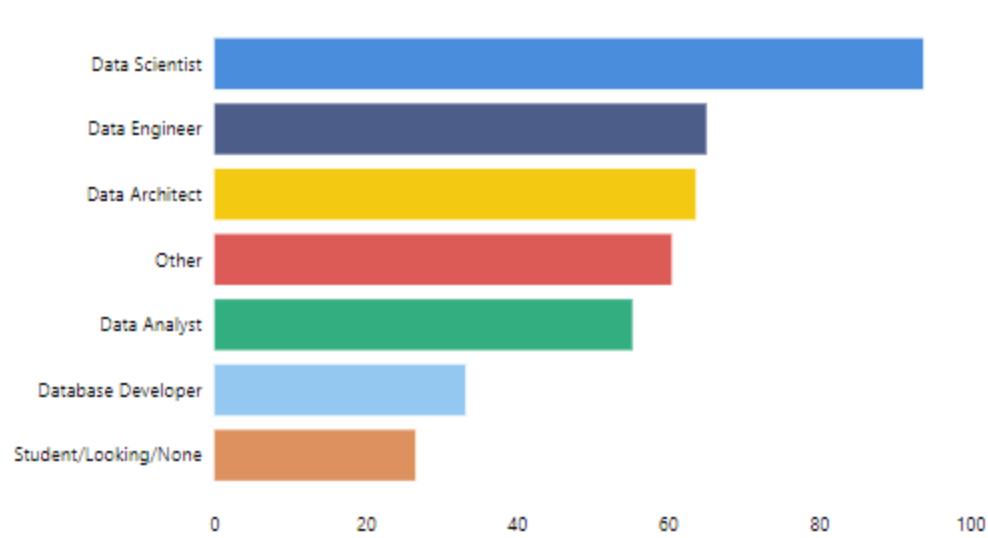
FIG 5: AVERAGE SALARY BY JOB TITLE (Designed with PowerBI)

The highest paid data professionals were Data Scientist with an average of $93,800; while the lowest paid were Database Developers, coming in at $33,200. Data Engineers received an average of $65,100; Data Architects came in at an average of $63,700; Data Analysts average salary stood at $55,300; while other fields in Data had an average of $60,500. With respect to the sectors, Finance had the highest average at $57,700 while Telecommunications had the lowest average at $40,300.
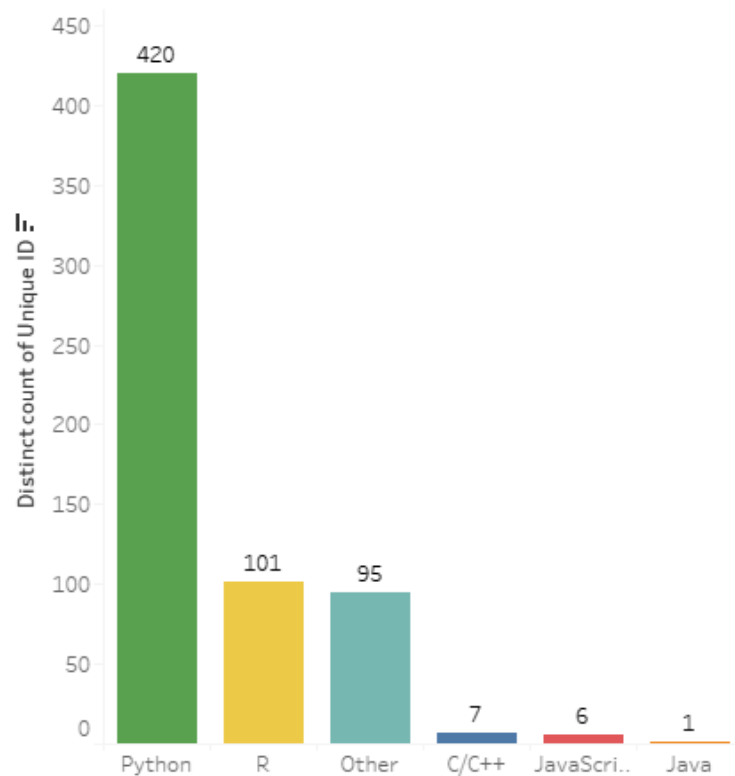
FIG 6: FAVORITE PROGRAMMING LANGUAGE (Designed with Tableau)

Python was selected as the most utilized programming language with a total of 420 respondents claiming that they preferred Python. R was the second favorite with 101 respondents; 6 people chose Javascript; 7 chose C/C++; 1 person chose Java and 95 people chose other languages.
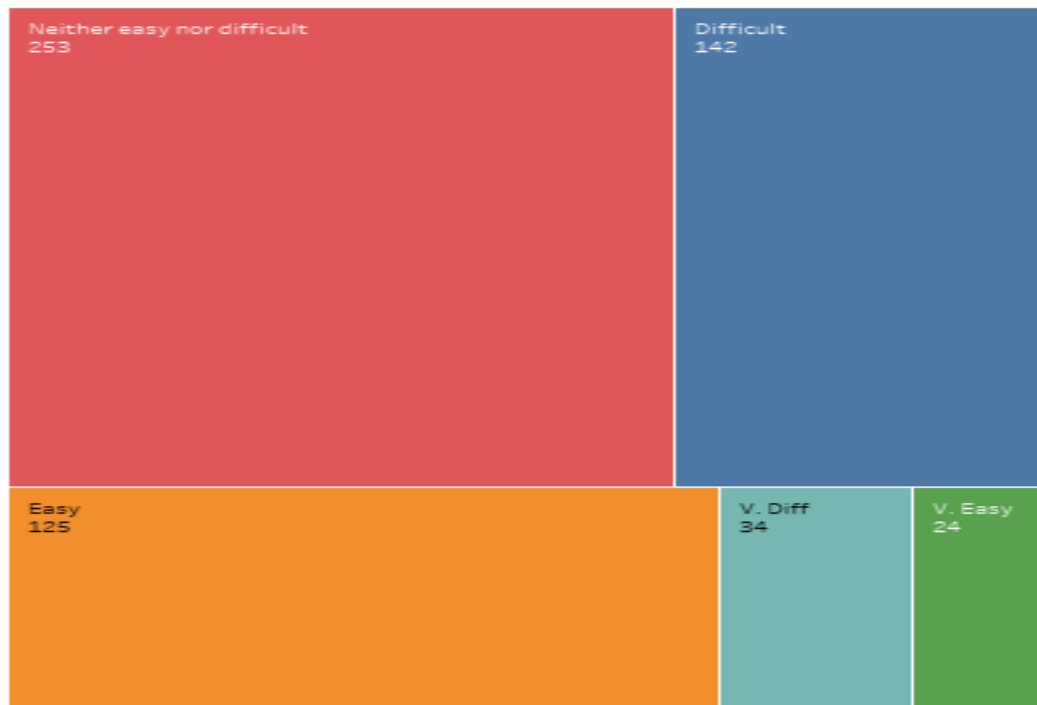
FIG 7: DIFFICULTY IN BREAKING INTO PROFESSION (Designed with Tableau)

The respondents were asked about the level of difficult they are experienced breaking into a career in data. 253 of them claimed it was neither easy nor difficult; 125 claimed it was easy; 24 claimed it was very easy; 142 responded that it was difficult and 34 maintained that it was very difficult.
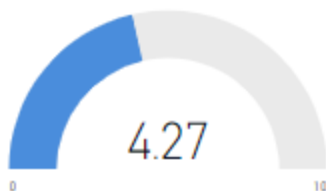

FIG 8: SATISFACTION WITH SALARY   **FIG 9: SATISFACTION WITH WORKLIFE BALANCE**

(Designed with PowerBI)

Various questions were asked to understand the level of job satisfaction with respect to certain areas. Satisfaction was graded on a scale of 1 to 10. With respect to salary, the analysis discovered that the average satisfaction was at an average of 4.27; and with respect to work life balance, the satisfaction score was 5.7

TABLEAU DASHBOARD
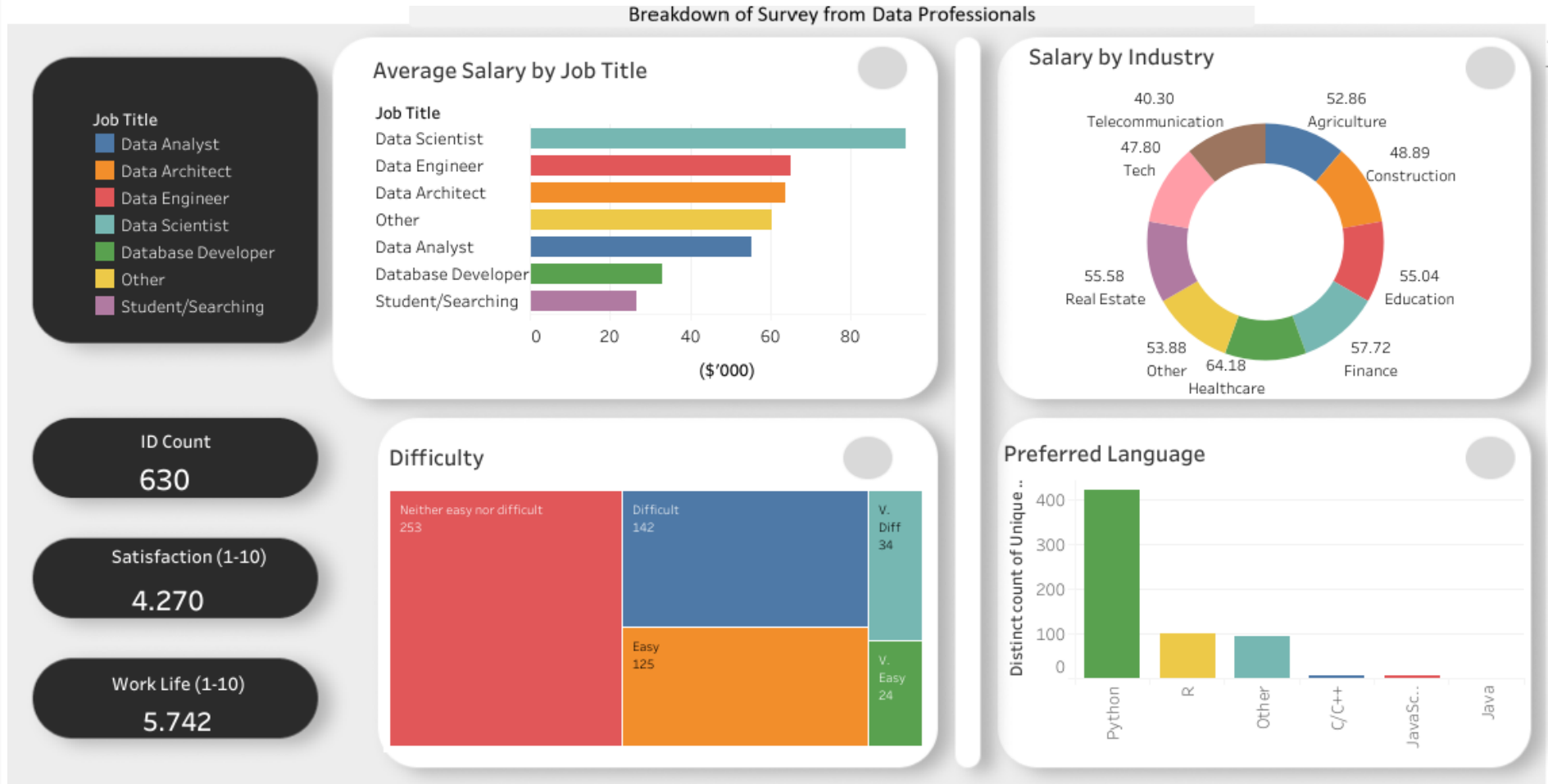
Click here to view the interactive Tableau dashboard

FIG 10: TABLEAU DASHBOARD

POWERBI DASHBOARD

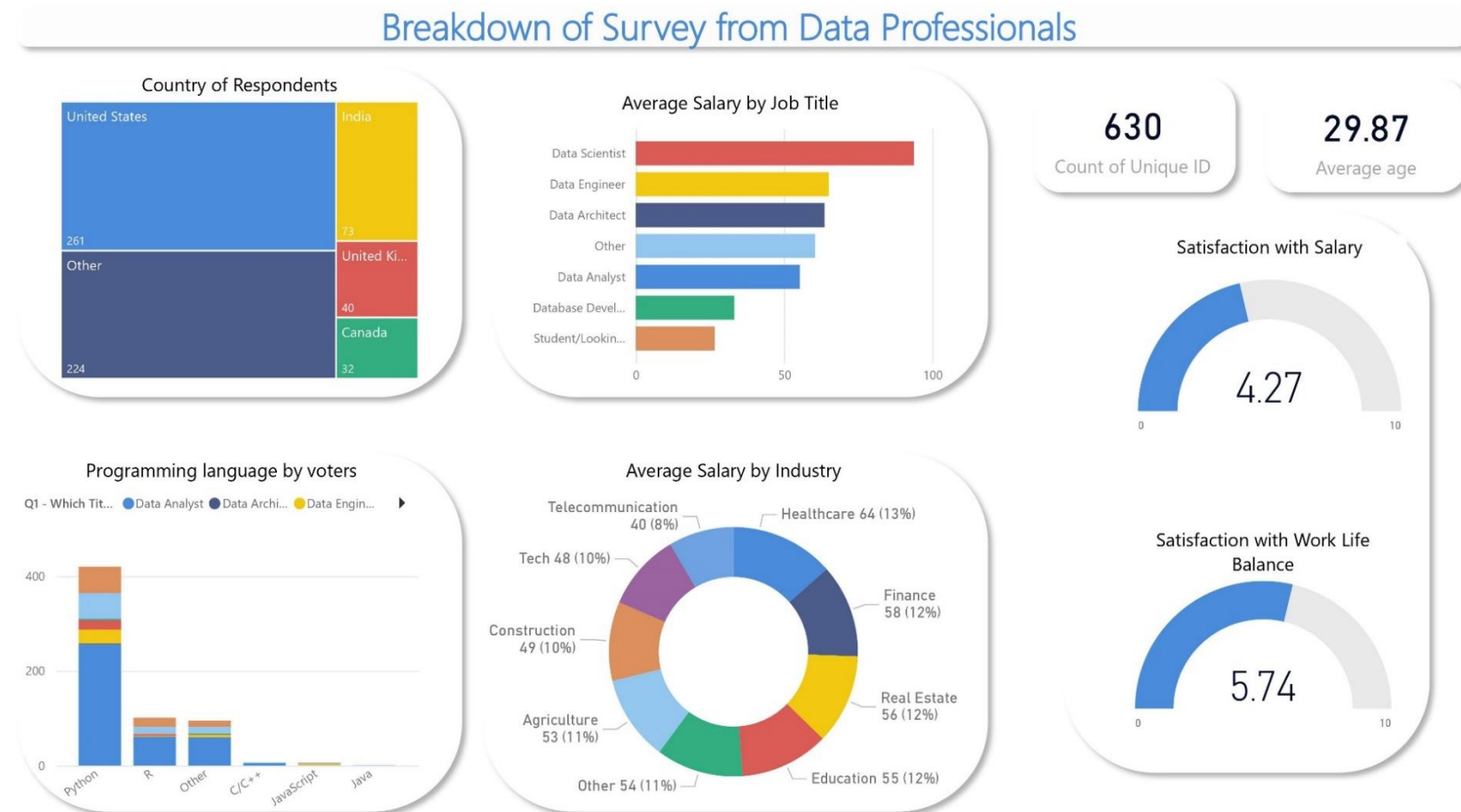Click here for to view the interactive Microsoft PowerBI Dashboard

FIG 11: POWERBI DASHBOARD

INTERPRETATION AND RECOMMENDATIONS

With a maximum salary of $225,000 it can be safely assumed that data professions are lucrative professions.

The analysis revealed that Data Scientists were the highest paid with an average salary of $93,800. However, given that there were only 5 Data Scientists, it may be difficult to make conclusive statements. Nevertheless, the reported average range is still not far off from other industry reports. The Finance industry also appears to be the most lucrative amongst the other industries. This means that for anyone considering transitioning into the profession with an eye on the attractiveness of the financial remuneration, a profession as a Data Scientist in the Finance industry would be the best option.

The statistics with respect to difficulty shows that transitioning into the field of data may not be as difficult as people think, with 72% of the responses ranging between 'very easy' and 'neither easy nor difficult', while only 28% claimed it was difficult.

It is however important to note that majority of the respondents were Data Analyst, and this could affect the absolute interpretation of the results. It is therefore advised that any interested person should interact with the dashboard to understand the level of difficulty for each field, alongside comparing the results of this survey with other industry reports.

With numerous programming languages out there, it could be difficult to get lost deciding which one would give the most returns for time invested. According to the analysis, Python came out as the overwhelming favorite of the data professionals with 420 respondents representing 67% preferring it over other languages, while R language come in at second with 101 people preferring it.

Therefore, aspiring and incumbent professionals who seek career growth should ensure they deepen their knowledge of Python. This would give them an edge in breaking into the field.

Despite the attractiveness of the industry, the satisfaction with salary was 4.27 out of a total score of 10. This indicates that majority of the professionals are not satisfied with their job. It could imply that they feel underpaid. However, a more detailed research would be needed to understand the actual cause of the dissatisfaction.

The work life balance also comes in at 5.7 which is neither good not bad. But still indicative of some level of dissatisfaction with work.

Given that this research has certain limitations, it is advised that the result of this analysis should not be taken as absolute. Further research should be carried out before making any career decisions.

LIMITATIONS

- The analysis was conducted by a Data Analyst. This may explain why there were more Data Analyst respondents.
- The analysis excluded SQL from the list of languages. Even though SQL is not classified as a programming language, it is still a much-used language in the field of data.
- The limited platforms utilized prevented the survey from reaching a larger pool of data professional.