

DATA PROFESSIONALS' SURVEY ANALYTICS.

SUMMARY

Career selection or career switch is not always easy as one needs to consider a number of factors before making the decision. Determinants such as ease/difficulty of the field, salary, work/life balance, team and management could affect the choice because of the physical and mental effects these have on work.

Data as the new oil is not exempt from this issue as there are many myths surrounding transition to data careers. To demystify this, a survey was done with sample size of about 630 data staff. As expected, this data will be dirty and need to be cleaned and transformed before any meaningful analysis for decision making is done.

The intent is to properly guide aspiring data enthusiasts on ease of transition, which area in data in terms of role, location to work at, work/life balance, management, team spirit. outlook to gender and favorite popular programming languages to make one relevant in the job market.

OBJECTIVES

To analyze the data professional survey and review determinants in the data field.

Understanding the data set and business problem

Examination of data set shows **28** columns with the most complete row of **630** elements.

All the data were in General format and needs to be converted into the respective data types such as numerical, date, currency etc.

The key columns are salary, current role, education, favorite programming language, industry, country of residence, and satisfaction with job, salary etc.

A few columns are completely empty and needs to be removed such as Browser, OS, City, country and referrer.

As expected, original dataset was dirty, with missing data, spelling errors, wrong data type assignment and needed a lot of cleaning to ensure reliable insights and decisions are made. There were no duplicates.

After the cleaning and transformation, 23 columns were remaining for analysis.

DATA CLEANING:

Data structure: After importing csv into Power Bi, Power Bi was able to assign the right data types to all the columns.

Removal of redundant columns: The anonymous, Browser, OS, City, country and referrer columns were removed by control, click on the columns> right click> remove columns

1. Splitting of current yearly salary to get average: This column is reported as K for thousand hence it was a string that cannot be used in meaningful analysis. To achieve this, >Right click column>split column> delimiter> dash. This converts the data to a form that can be converted to numerical data type. To complete this, replace the K. Choose both columns>replace values> K and nothing in the spaces. To convert to numerical, right click>change type> Whole number.

To get **average annual salary**: Add column>Custom column> Add both salary columns and divide by 2 to get average salary column.

2. To clean the question, “**In which country do you live?**”, the other category was further disaggregated to fish out individual countries. This was done to reduce the number of ‘Others’ as they were many and point out the specific country of residence. After using delimiters to split the column twice, both columns were merged by creating a custom column using:

let

```
primary = Text.Trim(Text.Upper([#"Q11 - Which Country do you live in?.1"])),
```

```
secondary = Text.Trim([#"Q11 - Which Country do you live in?.2.2"])
```

in

```
if primary = "OTHER" and secondary <> ""
```

```
then secondary
```

else Text.Trim([#"Q11 - Which Country do you live in?.1"]). The trim function was used to clear leading and trailing spaces while the upper function ensures that the data are in the same case format.

3. To clean up **education** column, replace function was used to replace the blanks with “No formal education”
4. To clean the **favorite programming language** column, After, using delimiter (:) to split the column, to ensure, that programming languages reported under “others” category are reduced. Then a custom column was created to merge the split programming language column, using:

```
if [#"Q5 - Favorite Programming Language.1"] = "Other"
```

```
then [#"Q5 - Favorite Programming Language.2"]
```

```
else [#"Q5 - Favorite Programming Language.1"].
```

The two columns programming columns were subsequently deleted.

5. To clean ‘**What industry do you work in?**’, using delimiters (‘ and ‘:’) to split the columns, the columns were merged into one by using a custom column syntax:
if Text.Trim(Text.Upper([#"Q4 - What Industry do you work in?.1"])) = "OTHER"
then Text.Trim([#"Q4 - What Industry do you work in?.2.2"])
else Text.Trim([#"Q4 - What Industry do you work in?.1"]) Trim and use of Upper text case was deployed to ensure that the spaces and text cases are taken care of as discussed above.
6. To clean “**What title best fits your current role?**’ After using delimiters twice to split the column, the columns were merged into one using a custom column:
= if Text.Trim(Text.Upper([#"Q1 - Which Title Best Fits your Current Role?.1"]))
)="OTHER"
then Text.Trim([#"Q1 - Which Title Best Fits your Current Role?.2.2"])
else Text.Trim([#"Q1 - Which Title Best Fits your Current Role?.1"])
7. **Average satisfaction:** A final column was created by adding up all the six columns tracking satisfaction and dividing each by 6 to give a weighted average of level of satisfaction. This was used in a slicer subsequently.

After these cleaning and transformation activities, the data was loaded into Power BI using the close and apply tab.

Data analysis

Creation of measures: A table of measures was created using enter data, then #Measures was used as a container for other measures. 3 measures namely **Average age, average salary and total number of respondents** were created.

Charts

9 Charts were created to track relationship between average salary and top ten countries of residence, industry, current title roles and highest education level.

Other charts traced the most common (5) programming languages, gender distribution of participants, how happy the respondents are with work/life balance and salary and finally top 4 determinants for new job search.

Slicers:

4 slicers were created in a dropdown form to show ethnicity, average satisfaction, ease/difficulty in breaking into data and career change into data.

Insights from the analysis

1. **Top Paying Industry:** The **Consulting industry** offers the highest average salary among data professionals.
2. **Highest Paying Role:** **Analytics Consultants and Directors** top the list of roles by average salary.
3. **Top Countries by Salary:** **UAE, Australia, and Ireland** lead in offering the highest average salaries.
4. **Education Impact:** A **PhD** leads to significantly higher average salaries than other degrees, with a steep drop from Master's to Bachelor's and below.
5. **Most Popular Programming Language:** **Python** is the dominant programming language, far ahead of others.
6. **Primary Job Switch Factor:** The biggest motivator for job search is **Better Salary**, followed by **Remote Work**.
7. **Gender Distribution:** There is a visible gender imbalance, with **males** making up the vast majority of respondents, though the average salary is higher in females.
8. **Work-Life Balance Satisfaction:** On a scale of 10, the average score is **5.74**, indicating room for improvement.
9. **Salary Satisfaction:** The score is **4.27**, showing **dissatisfaction with current pay**.
10. **Role vs. Salary Mismatch:** Some roles like **Software Engineers and Supply Chain Analysts** earn significantly lower than others—despite high demand—suggesting a pay gap by title or specialization.

Recommendations

1. **Upskill into Consulting:** Encourage mid-career professionals to pivot into consulting and strategic data roles for higher earnings.
2. **Geographic Job Strategy:** Promote **remote work or relocation** to top-paying countries (UAE, Australia) for better compensation.

3. **Promote Python Mastery:** Double down on **Python training**, especially for new entrants aiming for competitive advantage.
4. **Support Advanced Education:** Create **fellowship or sponsorship programs** to help professionals pursue Master's or PhD programs.
5. **Address Gender Gap:** Launch diversity hiring programs and mentorships to bridge the **gender imbalance** in data careers.
6. **Employer Salary Benchmarking:** Organizations should **benchmark salaries** against top industries and roles to remain competitive.
7. **Invest in Retention:** Improve **work-life balance and culture** as they're key drivers of attrition besides salary.
8. **Tailor Training for Low-Paid Roles:** Offer **specialized skill training** to those in low-paying roles (e.g., Supply Chain Analysts) to help them transition to higher-paying ones.
9. **Promote Remote Work Flexibility:** Since remote work is the second most desired feature, companies should offer flexible/hybrid models.
10. **Career Path Visibility:** Provide **transparent career ladders** showing how roles like Analyst or Engineer can evolve into Director/Consulting-level roles.