

# BUSINESS REPORT: DATA PROFESSIONAL SURVEY ANALYSIS

Name :SHASHANK KUMAR

MAJOR – BUSINESS ANALYTICS

REGNO. 2023JULB01112

## 1. PROBLEM STATEMENT

The field of data science and analytics has seen rapid growth over the past decade, with professionals transitioning into these roles from diverse educational and career backgrounds. This survey aimed to analyze responses from 630 participants to uncover key insights regarding their job roles, salary trends, work-life balance, salary satisfaction, and challenges faced when entering the industry.

The analysis seeks to answer the following critical questions:

- What are the most preferred programming languages among data professionals?
- How do salaries vary across different roles in the data field?
- How do professionals rate their work-life balance and salary satisfaction?
- What are the challenges faced by professionals while transitioning into data-related careers?

This study provides a comprehensive understanding of the data professional landscape to guide future workforce and organizational strategies.

## 2. DATA REQUIREMENTS

To derive meaningful insights, the following data elements were identified as essential:

---

### CORE SURVEY INFORMATION:

- **Job Roles and Titles:** To categorize participants into roles such as Data Scientist, Data Analyst, Data Engineer, etc.
- **Salary Data:** To assess current compensation trends across roles and industries.
- **Preferred Programming Languages:** To identify which languages are most popular and relevant in the data field.
- **Satisfaction Ratings:** To evaluate work-life balance, salary satisfaction, and upward mobility.

---

### DEMOGRAPHICS:

- **Age:** To analyze trends across different career stages.
- **Country:** To understand geographical influences on salaries and satisfaction.
- **Gender and Ethnicity:** To explore diversity in the data profession.

---

### CHALLENGES:

- **Difficulty of Entry:** To gauge the barriers faced by professionals entering the data industry

### 3. DATA COLLECTION

The data was collected through an online survey distributed via open-source platforms. Participants were asked to provide detailed responses to questions covering their job roles, experiences, challenges, and personal preferences.

#### DATA VOLUME:

- Total responses: **630 unique participants**
- Key demographic coverage: Countries, age groups, and gender representation.

#### DATA CONTEXT:

The survey was conducted as part of a project aiming to explore the career journeys and satisfaction levels of data professionals. The questions were designed to provide both quantitative and qualitative insights.

Unique ID	Email	Date Taken (America/New_York)	Time Taken (America/New_York)	Time Spent	Q1 - Which Title Best Fits your Current Role?.1	Q2 - Did you switch careers into
62a33b3db4da29969c62df3d	anonymous	06 October 2022	08:38:00	00:00:44	Data Analyst	Yes
62a33ce918134ddc75ce8c30	anonymous	06 October 2022	08:45:00	00:01:29	Data Analyst	Yes
62a33cf30a77c1a77f65baa2	anonymous	06 October 2022	08:45:00	00:02:31	Data Analyst	Yes
62a33d5c0f8c8599d5af107c	anonymous	06 October 2022	08:47:00	00:00:47	Data Analyst	Yes
62a33e38f3072dd8926221d3	anonymous	06 October 2022	08:51:00	00:01:25	Data Analyst	Yes
62a340d62e54c9003e5324b4	anonymous	06 October 2022	09:02:00	00:00:56	Data Analyst	Yes
62a34272f8d5fccc22f7d15f	anonymous	06 October 2022	09:09:00	00:01:09	Data Analyst	No
62a3449f7dc029e2d66bf5d1	anonymous	06 October 2022	09:18:00	00:01:15	Data Analyst	Yes
62a3452818134ddc75ce9e1b	anonymous	06 October 2022	09:20:00	00:01:33	Data Analyst	Yes
62a3455bbc6861bf3176d57a	anonymous	06 October 2022	09:21:00	00:02:36	Data Analyst	Yes
62a345c47dc029e2d66bf80a	anonymous	06 October 2022	09:23:00	00:01:36	Data Analyst	Yes
62a346a95d257a28a30d8cb9	anonymous	06 October 2022	09:27:00	00:00:58	Data Analyst	No
62a346d90a77c1a77f65d08d	anonymous	06 October 2022	09:27:00	00:02:39	Data Analyst	No
62a34726f8d5fccc22f7dec7	anonymous	06 October 2022	09:29:00	00:01:14	Data Analyst	No
62a349160a77c1a77f65d4db	anonymous	06 October 2022	09:37:00	00:00:28	Data Analyst	Yes
62a349c7cec54911e9ca67c7	anonymous	06 October 2022	09:40:00	00:01:05	Data Analyst	Yes
62a34c66b4da29969c6305af	anonymous	06 October 2022	09:51:00	00:01:36	Data Analyst	Yes
62a34ff1f408bae018ed601d	anonymous	06 October 2022	10:06:00	00:01:16	Data Analyst	No
62a3505bd8c00e9f5419f8df	anonymous	06 October 2022	10:08:00	00:01:22	Data Analyst	Yes
62a35130f3072dd892624d9f	anonymous	06 October 2022	10:12:00	00:02:04	Data Analyst	Yes
62a35541f3072dd892625517	anonymous	06 October 2022	10:29:00	00:02:07	Data Analyst	Yes
62a35b13bc6861bf31770188	anonymous	06 October 2022	10:54:00	00:01:05	Data Analyst	No
62a35fd0d24a2626027386a8b	anonymous	06 October 2022	11:11:00	00:01:03	Data Analyst	No
62a35f825d257a28a30dc128	anonymous	06 October 2022	11:13:00	00:02:17	Data Analyst	No
62a35fa0f8f7561645856c35	anonymous	06 October 2022	11:13:00	00:01:39	Data Analyst	No
62a361cc575fa3a5ae88f908	anonymous	06 October 2022	11:22:00	00:00:55	Data Analyst	No
62a3684ebc6861bf31771f15	anonymous	06 October 2022	11:50:00	00:00:48	Data Analyst	No

# DATA VALIDATION

Ensuring data accuracy and consistency was a critical step in this analysis. The following measures were taken:

## 1. Review for Completeness:

- Missing or invalid responses (e.g., empty fields or nonsensical entries) were identified and removed.
- Certain free-text fields were consolidated into standard categories (e.g., job roles, programming languages).

## 2. Relevance Check:

- Irrelevant fields like "Browser," "OS," and "Referrer" were dropped, as they did not contribute to the analysis objectives.
- Fields like "Salary Range" were converted into numeric averages for better analysis.

## 3. Standardization:

- Responses for job titles, programming languages, and countries were cleaned to remove duplicates and align them into fewer, more meaningful categories.

Time Taken (America/New_York)	Browser	OS	City	Country	Referrer
11:46					
11:47					
11:48					
11:49					
11:49					
11:50					
11:51					
11:53					
11:54					
11:55					
11:55					

# 5. DATA CLEANING

The raw dataset underwent the following cleaning steps to ensure it was ready for visualization:

- **Column Removal:**  
Irrelevant fields, such as *City*, *Browser*, and *Referrer*, were excluded. These fields did not add value to the analysis and cluttered the dataset.
- **Salary Processing:**  
The "Current Yearly Salary" field was cleaned by separating numeric and non-numeric entries. Numeric values were averaged to calculate an "Average Salary" metric, categorized by roles.
- **Categorical Consolidation:**
  - Job Roles: Similar titles (e.g., Data Architect, Solutions Architect) were grouped into broader categories like "Data Architect."
  - Programming Languages: Varied spellings and entries for languages like Python and R were standardized.
  - Country: Countries with fewer responses were grouped under "Other."

## Before

Q3 - Current Yearly Salary (in USD)
106k-125k
41k-65k
0-40k
150k-225k
41k-65k
0-40k
0-40k
125k-150k
86k-105k
41k-65k
66k-85k
0-40k
0-40k
0-40k
41k-65k
41k-65k
0-40k
0-40k
41k-65k
0-40k
41k-65k

## After

Average Salary
116
76
20
53
20
53
76
20
116
20
53
76

# DASHBOARD CREATION

The cleaned data was imported into Power BI, where tiling and visualizations were designed to uncover meaningful insights.

---

## DASHBOARD CREATION STEPS:

1. **Data Import and Transformation:**

The cleaned dataset was loaded into Power BI, and calculated measures were created for key metrics such as average salary, satisfaction scores, and work-life balance.

2. **Visual Design:**

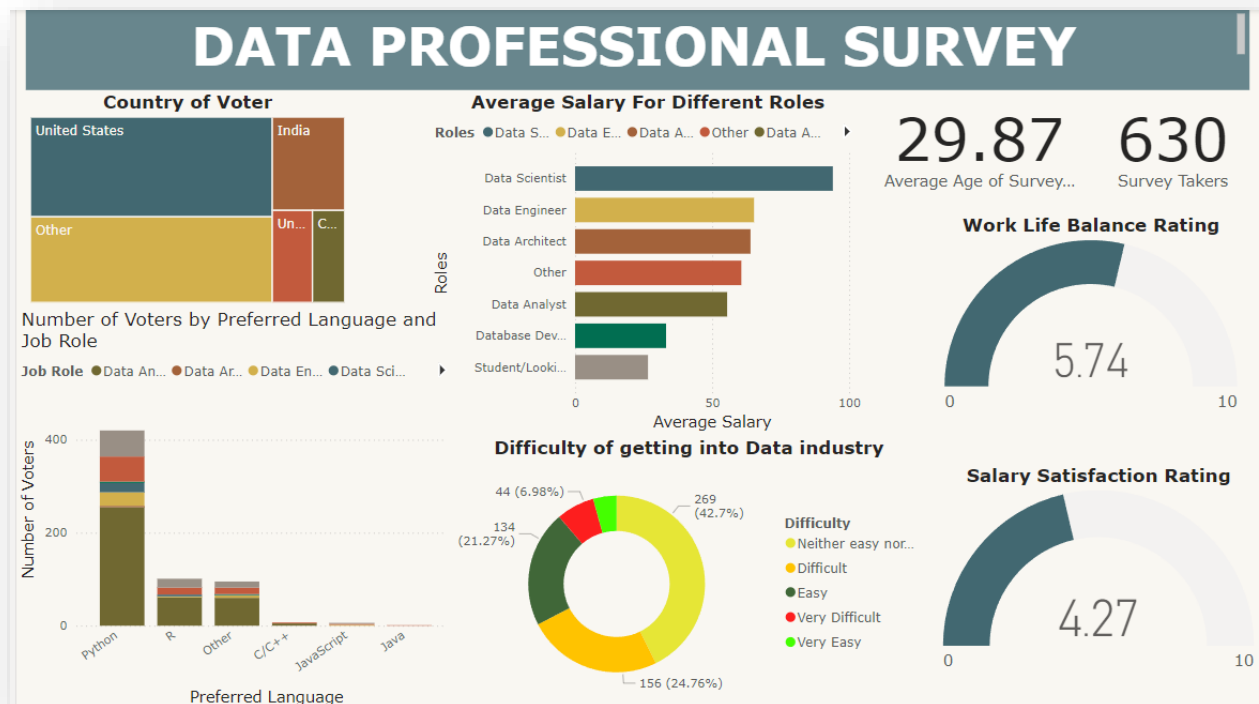
A dashboard was designed with interactive elements such as bar charts, pie charts, and KPIs. Each visualization was tailored to highlight specific insights:

- **Bar Charts:** Used to display average salaries for various job roles and the number of voters by programming language.
- **Pie Charts:** Represented difficulty ratings for breaking into the industry.
- **KPI Gauges:** Showcased satisfaction ratings for salary and work-life balance.

3. **Consolidation of Categories:**

The survey data was streamlined to ensure visuals were easy to interpret and insights could be drawn without overwhelming the audience.

## 7. FINAL DASHBOARD



The final Power BI dashboard summarized the survey data into actionable insights:

## KEY INSIGHTS:

### 1. Salaries by Roles:

- Data Scientists earned the highest average salary, followed by Data Engineers and Data Architects.
- Students or those looking to enter the industry had significantly lower average salaries.

### 2. Work-Life Balance:

- The overall work-life balance rating was moderate, with an average score of **5.74 out of 10**.
- This metric suggests there is room for improvement in providing better flexibility for professionals.

### 3. Salary Satisfaction:

- The salary satisfaction score averaged **4.27 out of 10**, indicating a potential gap between compensation expectations and reality.

### 4. Preferred Programming Languages:

- Python emerged as the most popular programming language, followed by R. Other languages such as C++ and JavaScript had significantly fewer votes.

### 5. Challenges in Breaking Into the Industry:

- A large percentage of respondents (42.7%) found it "Neither easy nor difficult" to enter the data field, while **21.3%** rated it "Difficult."

---

## VISUAL FEATURES:

- Dynamic filtering options allow users to explore insights by job roles, countries, or age groups.
- Clear, concise visuals make the data accessible to both technical and non-technical audiences.



## 8. STORYTELLING AND BUSINESS INSIGHTS

The survey results paint a comprehensive picture of the data professional landscape. Key takeaways include:

- **Salary Trends:** Organizations should benchmark salaries against industry standards to attract top talent, especially for high-demand roles like Data Scientist and Data Engineer.
- **Employee Satisfaction:** Moderate ratings for work-life balance and salary satisfaction highlight the need for better workplace policies and compensation structures.
- **Skills Development:** Python and R are the most valuable skills for aspiring data professionals. Training programs should emphasize these languages.
- **Industry Entry Challenges:** Companies and educational institutions can collaborate to reduce barriers for entry-level professionals by offering internships, mentorship, and clear career pathways.

By addressing these areas, businesses can improve employee retention, attract skilled professionals, and foster a more inclusive and satisfying work environment.