IBM Capstone Project

A Deep Dive into Technology Trends and Demographics

Presented by: Avik Sarkhel

Introduction

Overview:

- This presentation showcases insights from the IBM Capstone Project dataset, which contains 18,821 respondents.

- The original dataset had 114 columns and 18,845 rows, but after cleaning and transformation, 24 records with "Prefer not to say" age values were removed.

- The project was initially designed for Google Looker Studio, but this analysis was conducted using Power BI, modifying the provided questionnaire to enhance insights.

Purpose of the Analysis:

- Understand current technology usage and preferences in programming languages, databases, web frameworks, and platforms.

- Identify future technology adoption trends, helping professionals and businesses prepare for emerging shifts.

- Explore respondent demographics to understand how age, education, and geography influence technology adoption.

Project Overview

Dataset: Originally contained **114 columns & 18,845 rows**; cleaned dataset now has **18,821 respondents**.

Goal: To analyze technology usage trends and respondent demographics.

Methodology:

- **Data Cleaning:** Removed irrelevant columns, handled multi-value responses.

- **Data Transformation:** Split multi-value responses (e.g., languages, databases) using Power BI.

- **Dashboard Creation:** Designed **3 dashboards** focusing on **Current Usage, Future Trends, and Demographics**.

Data Preparation & Transformation

Data Loading: Imported dataset into Power BI.

Query Duplication: Created **3 separate queries**, each cleaned for specific dashboard needs.

Handling Multi-Value Fields:
- Columns like **Languages Used** and **Databases Used** contained multiple values separated by semicolons.
- Used **split by rows** to normalize data for accurate visualization.

Demographics Adjustments:
- Removed **24 records** where respondents marked "Prefer not to say" for age.
- Ensured consistency in country and education-level classification.

Questionaire and modifications to the original project structure

The original questionnaire provided a structured approach to analyzing technology usage, future trends, and demographics. To enhance clarity and improve insight extraction, modifications were made by adjusting visualization types and removing redundant elements. Below is a comparison of the original and modified questionnaire :

Category	Original Visualization	Modified Visualization
Current Technology Usage	• Top 10 Languages Used (Stacked Bar Chart) • Top 10 Databases Used (Stacked Column Chart) • Top 10 Platforms Used (Word Cloud) • Top 10 Web Frameworks Used (Scatter Bubble)	• Top 10 Languages Used (Horizontal Lollipop Chart) • Top 10 Databases Used (Vertical Lollipop Chart) • Top 10 Platforms Used (Funnel Chart) • Top 10 Web Frameworks Used (Pie Chart)
Future Technology Trends	• Top 10 Languages Used (Stacked Bar Chart) • Top 10 Databases Used (Stacked Column Chart) • Top 10 Platforms Used (Tree map) • Top 10 Web Frameworks Used (Scatter Bubble)	• Top 10 Languages Desired (Horizontal Lollipop Chart) • Top 10 Databases Desired (Vertical Lollipop Chart) • Top 10 Platforms Desired (Funnel Chart) • Top 10 Web Frameworks Desired (Pie Chart)
Demographics	• Respondents by Age (Pie Chart) • Respondent Count by Country (Map Chart) • Respondent Distribution by Education (Line Bar Chart) • Respondent Count by Age, Classified by Education Level (Stacked Bar Chart)	• Respondents by Age (Horizontal Lollipop Chart) • Respondent Count by Country (Map Chart) • Respondent Distribution by Education (Vertical Lollipop Chart) • Removed (Covered by above visuals)

Data Visualisation
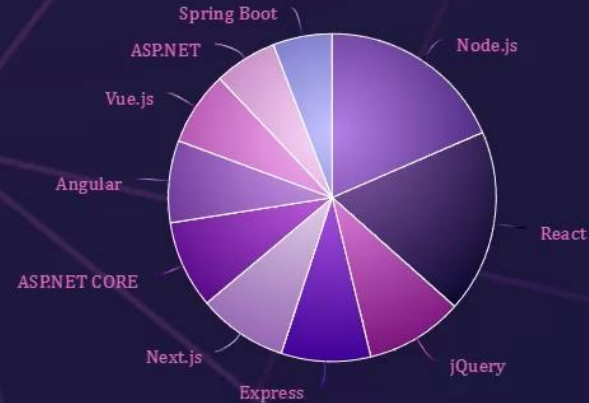

FUTURE TECHNOLOGY USAGE

Top 10 Database Desired
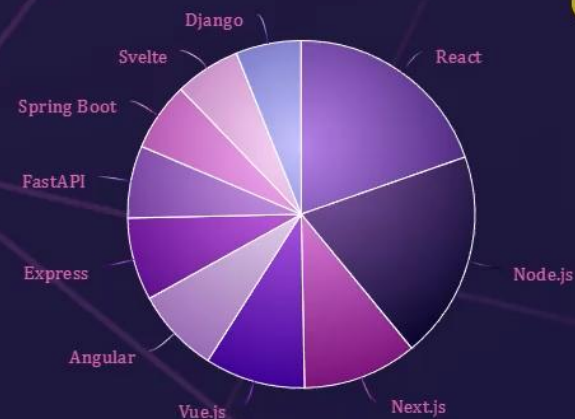
Database Desired	Number of Respondents
PostgreSQL	12,193
Redis	6,384
SQLite	6,295
MySQL	6,204
MongoDB	5,618
Microsoft SQL Server	4,345
Elasticsear...	3,665
MariaDB	3,078
Dynamodb	2,154
Supabase	1,623

Top 10 Platform Desired

Platform	
Amazon Web Services (A...	100%
Microsoft Azure	61%
Google Cloud	52%
Cloudflare	40%
Digital Ocean	28%
Vercel	26%
Firebase	26%
Hetzner	15%
Supabase	13%
Netlify	12%

Top 10 Language Desired

Language Desired	Number of Respondents
JavaScript	11
SQL	10,944
TypeScript	10,437
HTML/CSS	10,016
Python	8,919
Go	5,661
Rust	5,597
C#	5,590
Bash/Shell (all s...	5,582
Java	4,048

Top 10 Webframe Desired

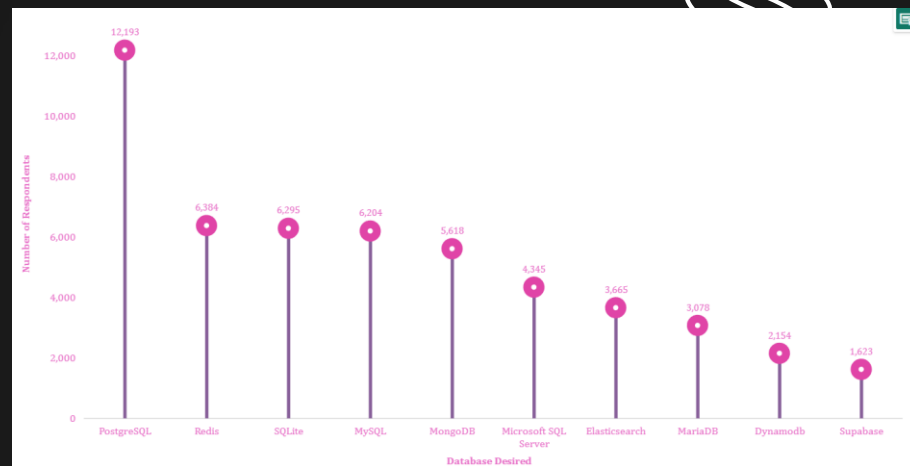Django, Svelte, Spring Boot, FastAPI, Express, Angular, Vue.js, Next.js, Node.js, React
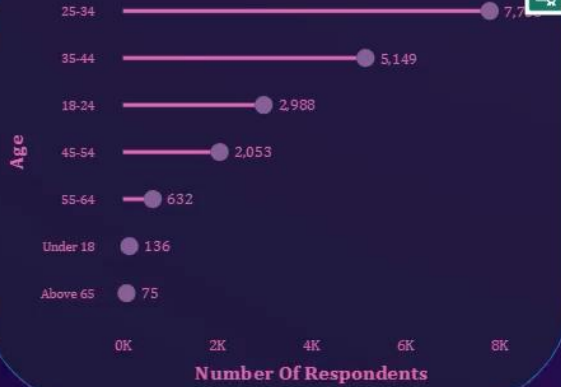
Comparison of Current and Desired Database Usage

Analysis:

- **PostgreSQL** remains the most used and most desired database, showing its strong preference among developers.
- **MySQL**, currently the second most used, sees a lower ranking in desired databases, suggesting a potential shift in preference.
- **SQLite** and **Redis** gain popularity in future preferences, indicating their growing importance in development.
- **MongoDB** and **Microsoft SQL Server** remain stable in both current and desired rankings.
- **Supabase** appears in the desired list but not in the current top 10, indicating emerging interest.

Comparison of Current and Desired Platform Usage

Left funnel (current usage):
- Amazon Web Services (AWS) — 26%
- Microsoft Azure — 16%
- Google Cloud — 13%
- Cloudflare — 9%
- Digital Ocean — 8%
- Firebase — 7%
- Vercel — 7%
- Heroku — 5%
- Netlify — 4%
- Hetzner — 3%

*This view is only for preview purpose only. Actual may vary based on the size of visual.

Right funnel (desired usage):
- Amazon Web Services (AWS) — 100%
- Microsoft Azure — 61%
- Google Cloud — 52%
- Cloudflare — 40%
- Digital Ocean — 28%
- Vercel — 26%
- Firebase — 26%
- Hetzner — 15%
- Supabase — 13%
- Netlify — 12%

*This view is only for preview purpose only. Actual may vary based on the size of visual.

Analysis:

- **Amazon Web Services (AWS)** remains the dominant platform in both current and desired usage, reinforcing its industry leadership.
- **Microsoft Azure** and **Google Cloud** hold strong positions but see a shift in ranking, suggesting evolving preferences among developers.
- **Cloudflare** maintains its relevance, but its future desirability is slightly lower.
- **Supabase** emerges in the desired platforms list, indicating increasing interest in modern backend solutions.
- **DigitalOcean** and **Hetzner** gain future traction, hinting at growing demand for cost-effective cloud solutions.

# Comparison of Current vs. Desired Programming Languages
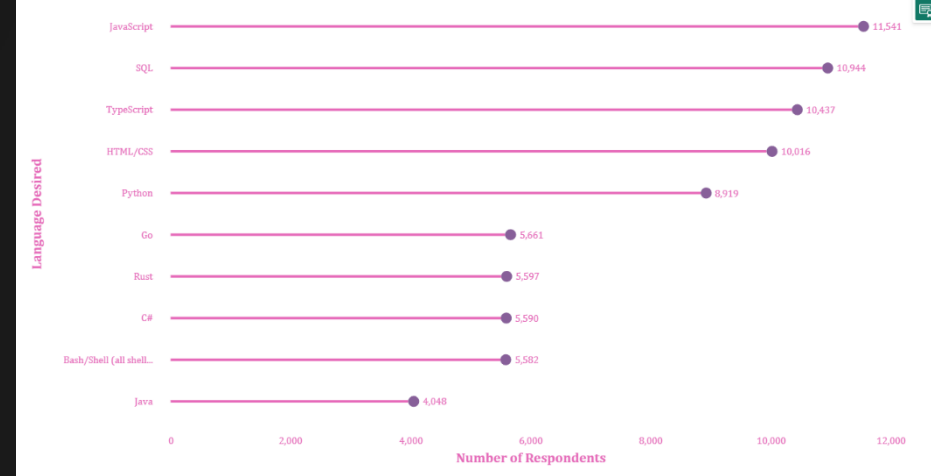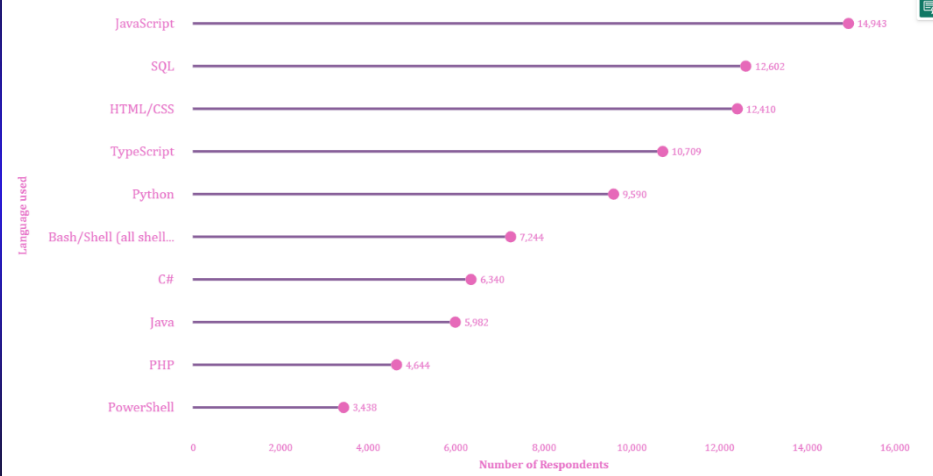
Analysis:

- **JavaScript** remains the most used and desired language, reflecting its dominance in web development.
- **SQL** sees a slight drop in desirability compared to current usage, indicating a potential shift in database management trends.
- **TypeScript** and **HTML/CSS** remain highly relevant but show a moderate decline in future demand.
- **Go** and **Rust** appear in the desired languages list, highlighting growing interest in modern, high-performance languages.
- **C# and Java** see relatively stable demand, but Bash/Shell scripting experiences a slight decline in future preference.

# Comparison of Current vs. Desired Web Frame
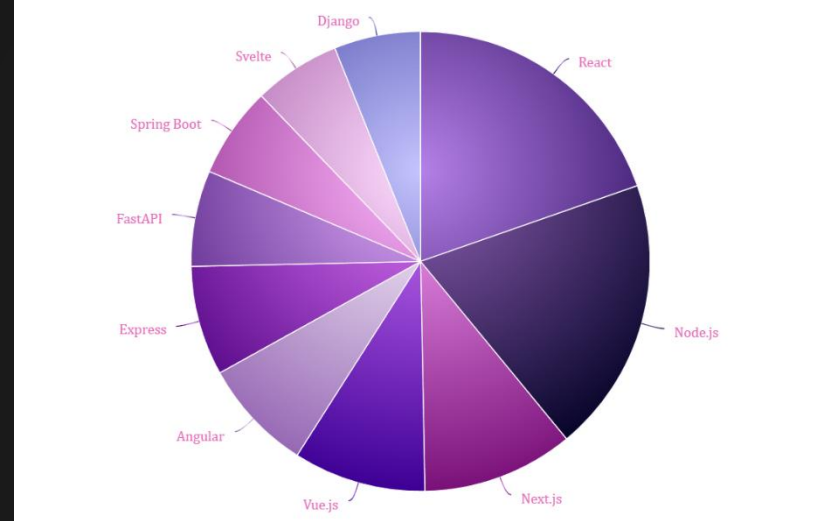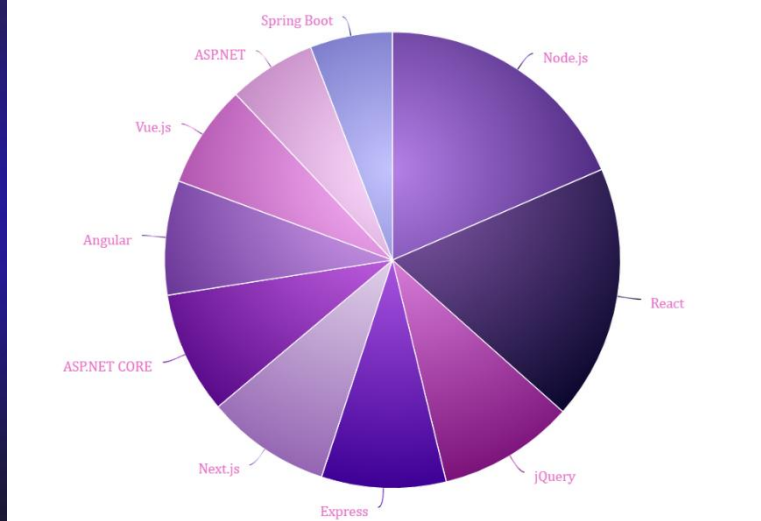
Analysis:
- **React and Node.js** remain dominant in both current and desired frameworks, showing their continued relevance.
- **ASP.NET and jQuery** are present in current frameworks but are absent in the desired ones, indicating a shift away from older technologies.
- **Next.js** and **Vue.js** continue to be popular choices, appearing in both lists.
- **FastAPI, Django, and Svelte** emerge in the desired frameworks list, signaling increased interest in modern, lightweight, and efficient web frameworks.
- **Express and Spring Boot** maintain their presence, showing their reliability for backend development.

Age Distribution of Respondents
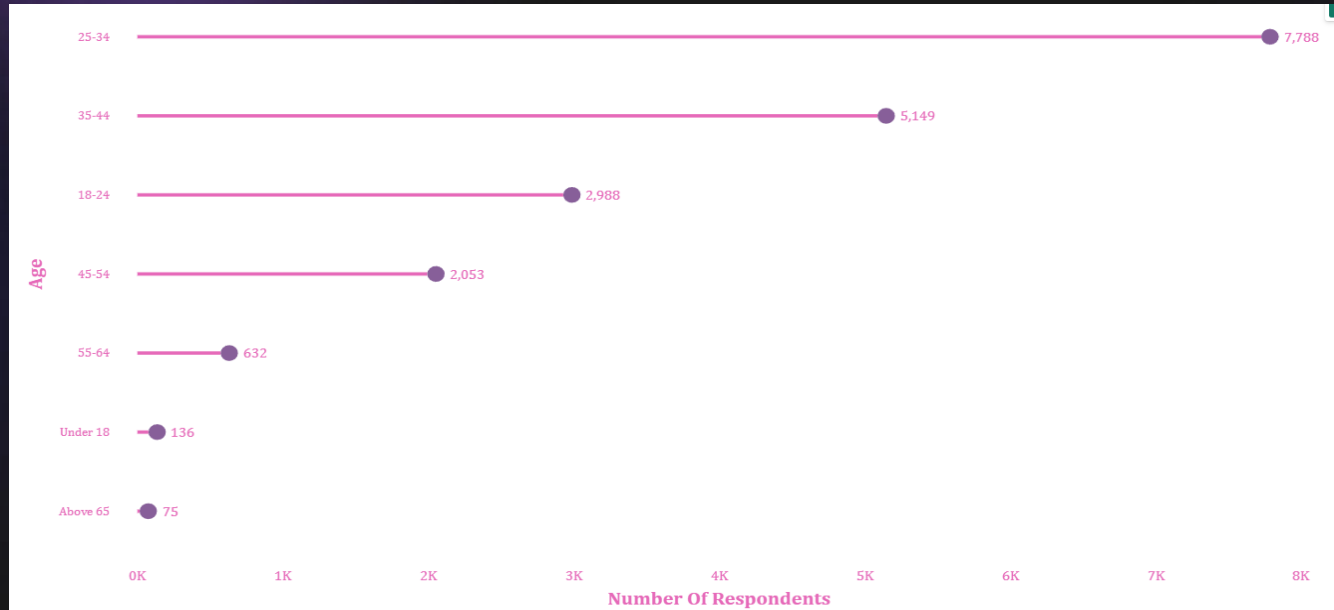
Analysis:

- **25-34 age group dominates** with **7,788** respondents, making it the largest segment.
- **35-44 age group** follows with **5,149** respondents, showing a significant presence.
- **18-24 age group** has **2,988** respondents, indicating strong participation from young adults.
- **45-54 group** has a moderate representation with **2,053** respondents.
- **Participation drops significantly** for older age groups:
- **55-64**: **632 respondents**
- **Above 65**: **75 respondents**
- **Minimal participation from Under-18** with just **136 respondents**.

Geographic Distribution of Respondents
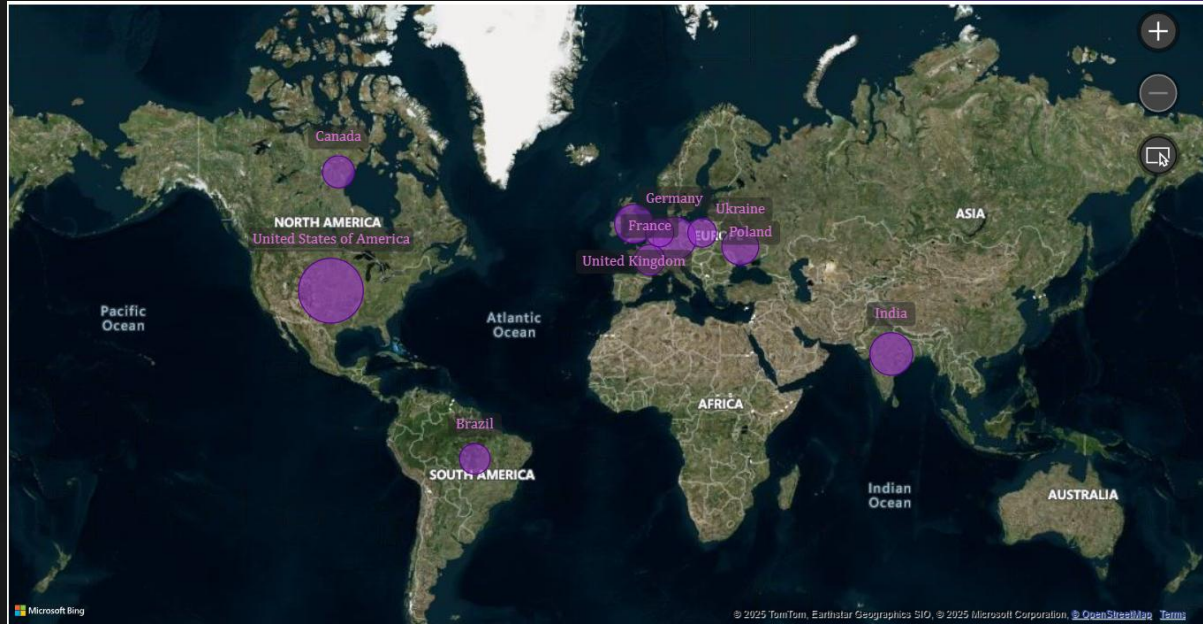
Analysis:

• **United States** has the highest number of respondents (**3,433**), making up the largest segment.

• **Germany (1,338)** and **India (1,315)** follow as the next largest contributors.

• **United Kingdom (1,052)** and **Ukraine (931)** show significant participation.

• **Canada (659), France (598), and Brazil (587)** contribute moderately.

• **Poland (494) and Netherlands (479)** round out the top 10.

Educational Background of Respondents
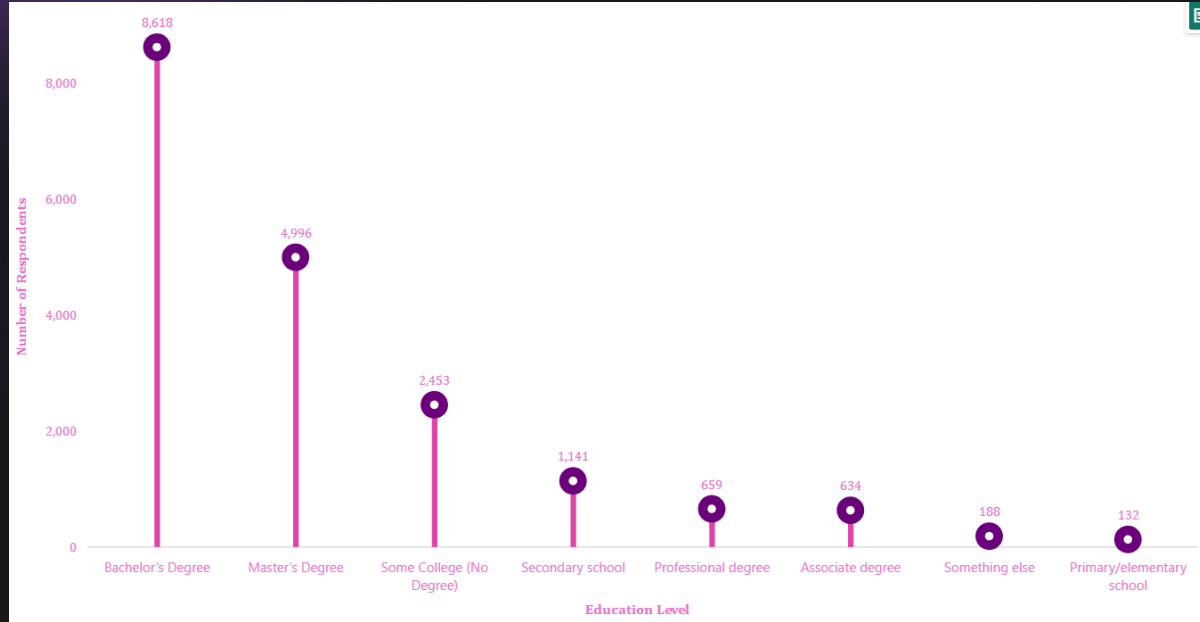
Analysis:

- **Bachelor's Degree holders (8,618)** form the largest group, indicating a high level of formal education.
- **Master's Degree (4,996)** is the second most common, showing significant postgraduate education.
- **Some College, No Degree (2,453)** and **Secondary School (1,141)** represent respondents with incomplete higher education.
- **Professional Degrees (659)** and **Associate Degrees (634)** have relatively smaller representation.
- **Primary/Elementary School (132)** and **Something Else (188)** indicate a minimal number of respondents with alternative education paths

Conclusions

In this Project a dataset of **18,821 respondents were** analyzed to explore **technology usage, future adoption trends, and demographics**. The study focused on identifying the most used and desired **programming languages, databases, cloud platforms, and web frameworks** while also analyzing **age distribution, geographic representation, and educational backgrounds**. The data was transformed and visualized using **Power BI**, improving clarity and insight extraction.

The **Key findings** found from this project are:

- **Most used languages**: JavaScript, Python, SQL; **emerging**: Go, Rust.
- **Top database**: PostgreSQL; **rising interest in**: Supabase.
- **Cloud leader**: AWS; **growing interest in** DigitalOcean, Hetzner.
- **Web frameworks**: React, Node.js dominate; jQuery **declining**.
- **Major age group**: 25-34 years.
- **Top countries**: U.S., Germany, India.
- **Common education levels**: Bachelor's & Master's degrees.

These insights help businesses and organizations identify future technology trends, refine hiring strategies, and make data-driven decisions on tech adoption.

Project Resources & Contact Information

GitHub Repository: (CLICK HERE) *(For project files & code)*
Google Drive Link: (CLICK HERE) *(For large files that couldn't be uploaded to GitHub)*
Coursera Project Link: (CLICK HERE)   *(For reference to the original capstone project)*

THANK YOU !!!!

Do you have any questions?

Write me at avik305sarkhel@gmail.com