IBM DATA ANALYST CAPSTONE PROJECT 2025

BY:

BIMAL UPADHAYA

24/07/2025

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Data Visualisation
  - Dashboard
- Discussion
  - Findings and Implications
- Conclusion
- Appendix
- Links

EXECUTIVE SUMMARY

1. Current Technology Landscape: The developer community predominantly uses JavaScript, Python, and SQL, with PostgreSQL and MySQL leading in database usage, AWS, Azure, and Google Cloud as top platforms, and React and Node.js as key web frameworks.
2. Future Technology Trends: Anticipated future demand remains focused on web, cloud, and data-centric technologies, with JavaScript, SQL, TypeScript, PostgreSQL, Redis, and Node.js/React being the most desired skills.
3. Demographic Profile: Respondents are primarily young to middle-aged (25-44 years old), highly educated with Bachelor's or Master's degrees, and largely represented by the United States, India, and Germany.
4. Compensation Trends: Yearly compensation strongly correlates with both age and professional coding experience, with the 55-64 age group earning the highest compensation.
5. Job Satisfaction & Professional Role: Most participants are professional developers who report high overall job satisfaction (8 out of 10), with the 55-64 age group also achieving the highest satisfaction rates alongside top compensation.

INTRODUCTION

- **Purpose**

  This report provide a data driven understanding of the current and future trend within the developer ecosystem which includes technology adoption, demographic profiles, compensation benchmarks and job satisfaction dynamics.

- **Target Audience**

  This report is primarily intended for Tech Company Leadership, Individual Developers and Educational Institutions, however anyone with the foundational understanding of data can get insight about the scenario.

- **Value Provided**

  1. This report can inform talent acquisition and development strategies by identifying key demographic segments and their current and future skill preferences.
  2. The report also offer an insight into compensation trends by age and experience. Also, offers a data driven information about emerging technologies which helps in career guidance for the individuals who are interested in tech development.
  3. Moreover, this report enable the organisations to benchmark their internal practices against industry trends in technology usage, compensation and employee satisfaction.

METHODOLOGY

- **Data Source:** The analysis is done using the Stack Overflow Survey, cleaned and organised by IBM teams and provided in the form of URL so that we can fetch or download.
- **Data Load, Preprocessing and Cleaning:** Data is first being load into the python library pandas data frame as csv file and checked it's structure. After that, removed or filled the null values, remove the duplicates and unnecessary columns for faster processing, find the outliers and used the log transform technique to balance the data for column like ConvertedCompYearly. Used the binning, grouping, ordering and mapping techniques for categorical data like age for correct visualisation. Used the explode and split function for the rows which has multiple data like LanguageWantToWorkWith.
- **Data Analysis Tools:** Python (Pandas and NumPy for data manipulation, Matplotlib and Seaborn for visualisation) and SQLite for querying the data. For Statistical Measures, calculated the Median values for compensation and job satisfaction.
- **Visualisation:** Different visualisation charts across the analysis using Google Looker Studio and Python Libraries(Seaborn and Matplotlib). Line Chart, Bar Chart, Pie Chart, Map Chart, Bubble Chart, Box Plot, Tree Map and World cloud are some data visualisation chart used in the analysis process.

PROGRAMMING LANGUAGE TRENDS

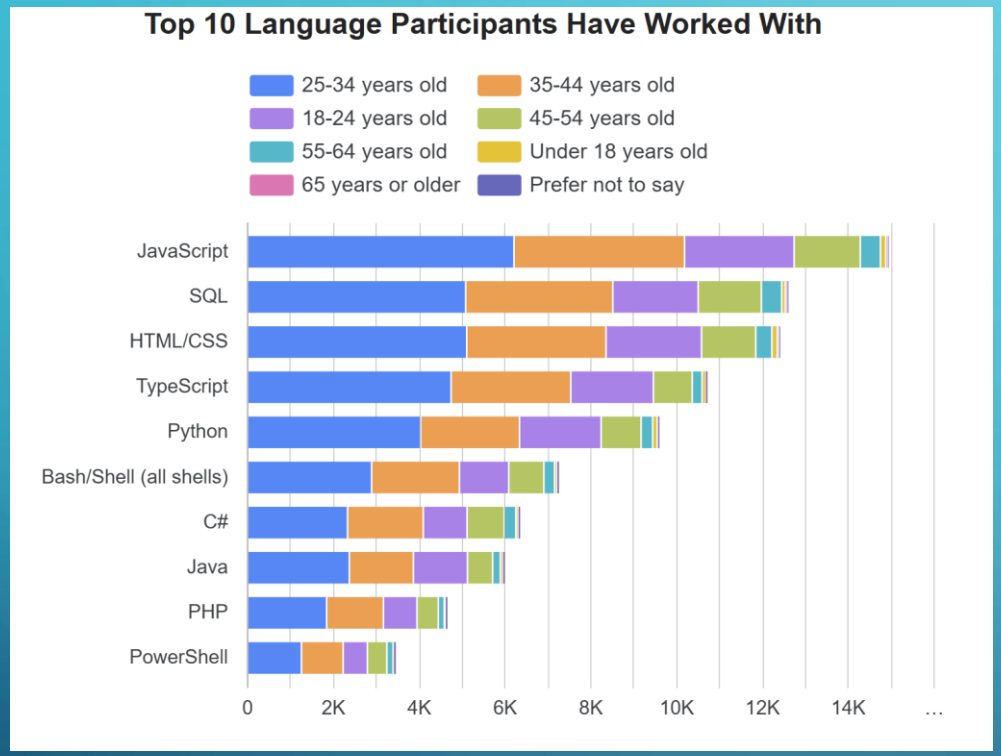

Fig. 1: Current Language Usage

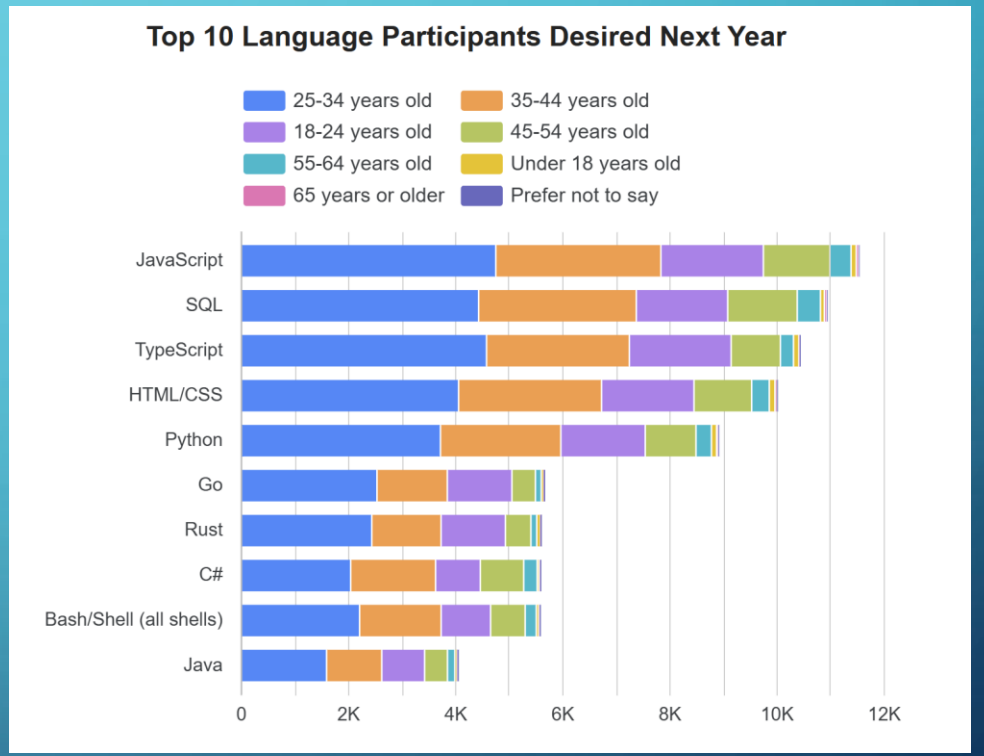

Fig. 2: Future Language Trend

FINDINGS AND IMPLICATIONS FOR LANGUAGE

- Finding
  1. For the current year, JavaScript, SQL and HTML/CSS are the widely used language along with Python by the participants. The largest segments of users across most used languages is the age groups in between 25-34 and 35-44.
  2. For the next coming years, participants would love to work with JavaScript, SQL, TypeScript and Python. Notably, while not used widely, languages like Go and Rust appear higher on the desired list. The desired for new language is also largely driven by the age group between 25-34 and 35-44.
- Implication
  1. For current year, languages like JavaScript, SQL, HTML/CSS and Python represent the current core skillset in the industry and companies should prioritise hiring and training in there areas for existing projects.
  2. For future, the trends seems to change in terms of language usage. Businesses should consider investing in training and exploring projects with Go and Rust to stay competitive, while continuing to support and develop expertise in JavaScript, SQL and Python. Learners also need to consider the emerging languages to stay competitive in the job market. Furthermore, Educational institutions should align curricula with these desired future skills.

DATABASE TRENDS

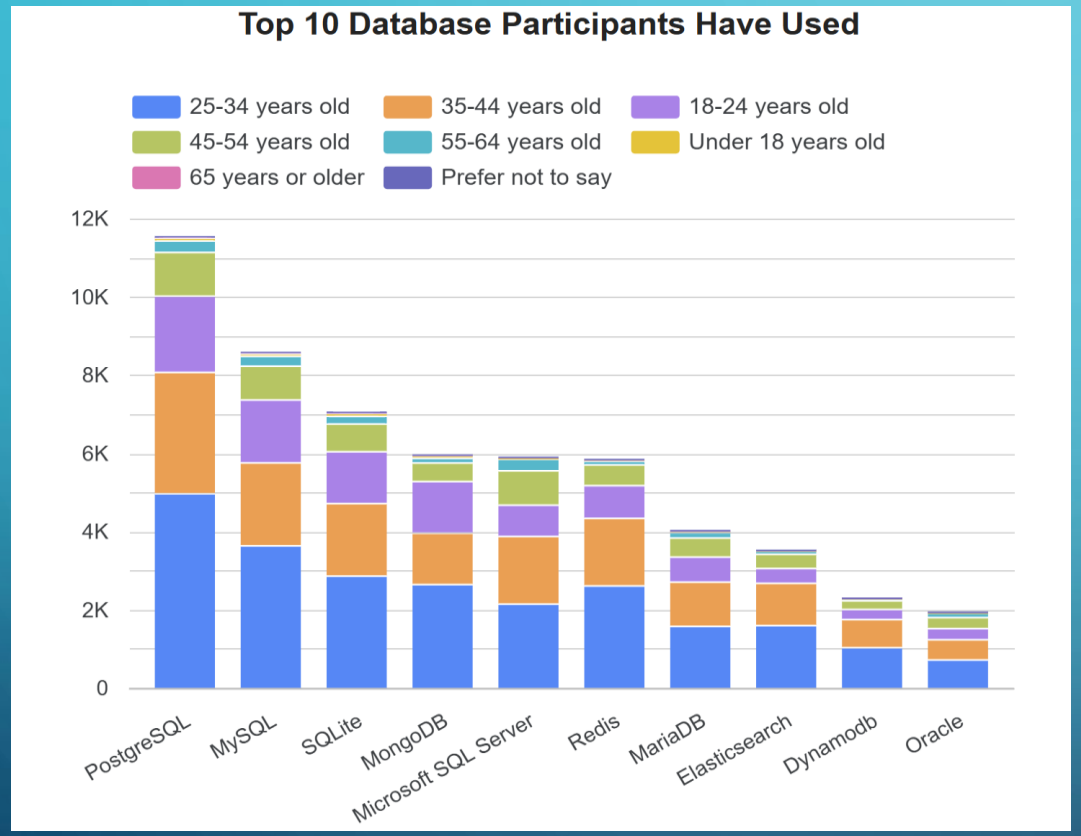

Fig. 3: Current Database Usage

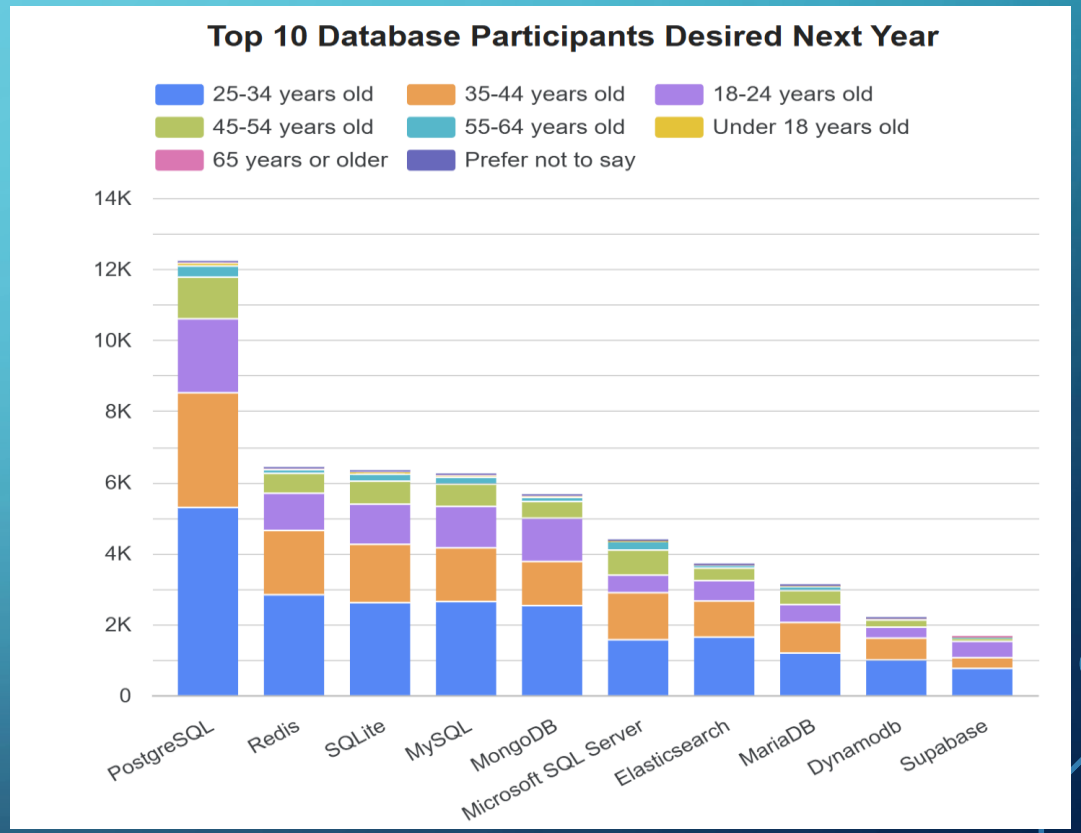

Fig. 4: Future Database Trends

FINDINGS AND IMPLICATIONS FOR DATABASE

- Finding
  1. For the current year, PostgreSQL is the most used database by the participants followed by MySQL. Databases like SQLite, MongoDB and Microsoft SQL Server also show the significant current usage. The 25-34 and 35-44 age groups is the largest user base for most of the database.
  2. For the next coming years, PostgreSQL maintains its lead as most desired database for future work. On the other hand, Redis and SQLite show a notable increase in desire, ranking higher than their current usage. Same as current database usage, the age group between 25-34 and 35-44 has the primary divers of interest in future database technologies.
- Implication
  1. For current year, PostgreSQL, MySQL and SQLite represents the foundational data storage technologies implemented in current projects and expertise in these database remains crucial for working in existing systems.
  2. For the future trend, the strong and continued desire for PostgreSQL reinforce its position as a go-to database for new developments and modern architectures. On the other side, rising preference for Redis suggests a growing need for high performance storage system. Also, Increased interest can be seen from the visualisation in SQLite might point to a demand for lightweight, embedded database solution for different applications.

BAR PLOT OF JOB POSTING

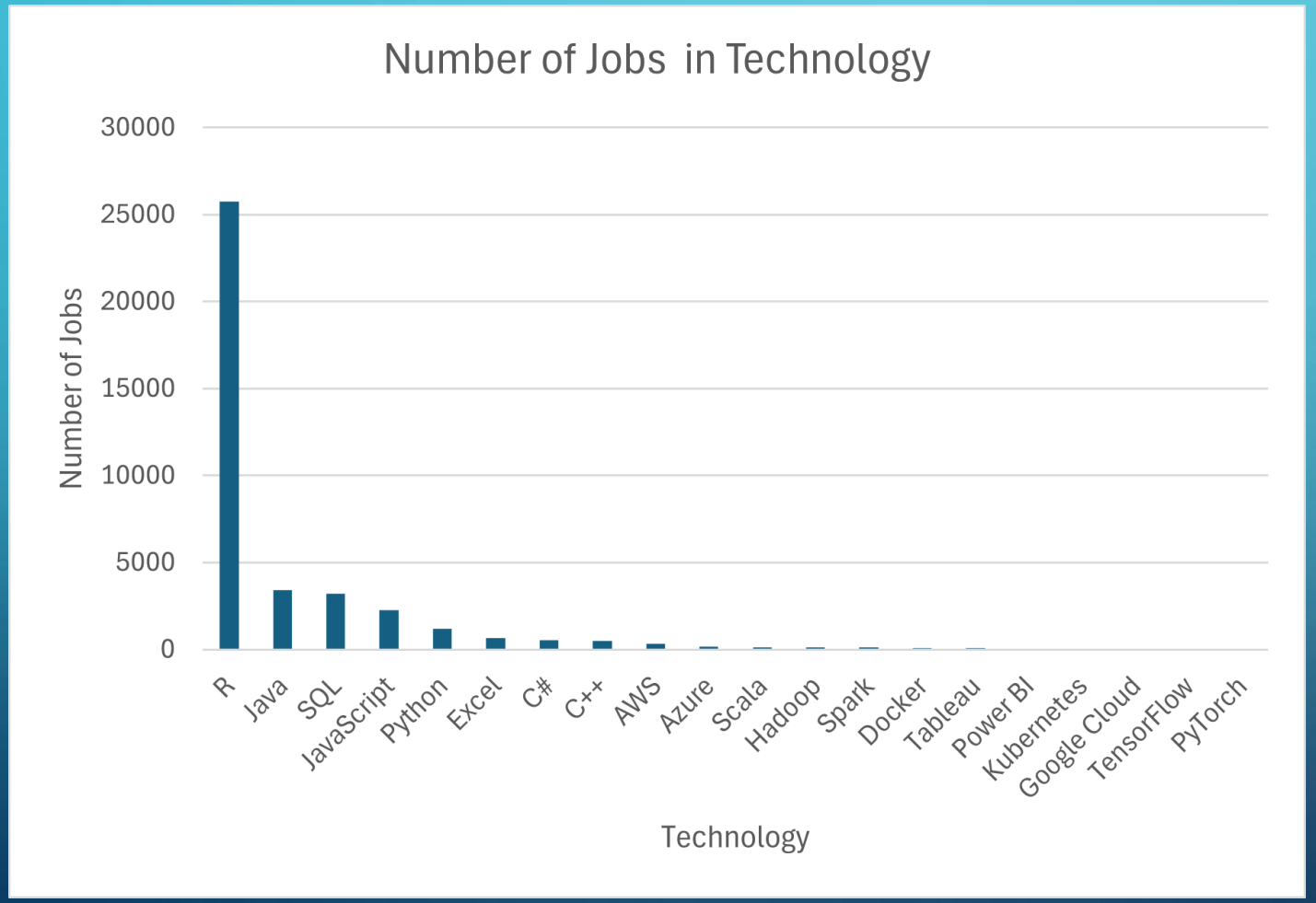

Fig. 5: Bar chart of Job posting

The figure illustrates the number of jobs posting in different sector of technology.

- Language R has the highest number of job posting with around 26000, which shows the strong demand of individuals who knows R in the job market.
- Java comes second with around 4500 followed by SQL with almost the equal number of job posting.
- SQL, JavaScript and Python can also be seen in the list of job posting.
- Jobs in Excel, C#, C++, AWS and Azure can be seen with small number of opening.

BAR CHART OF POPULAR LANGUAGES

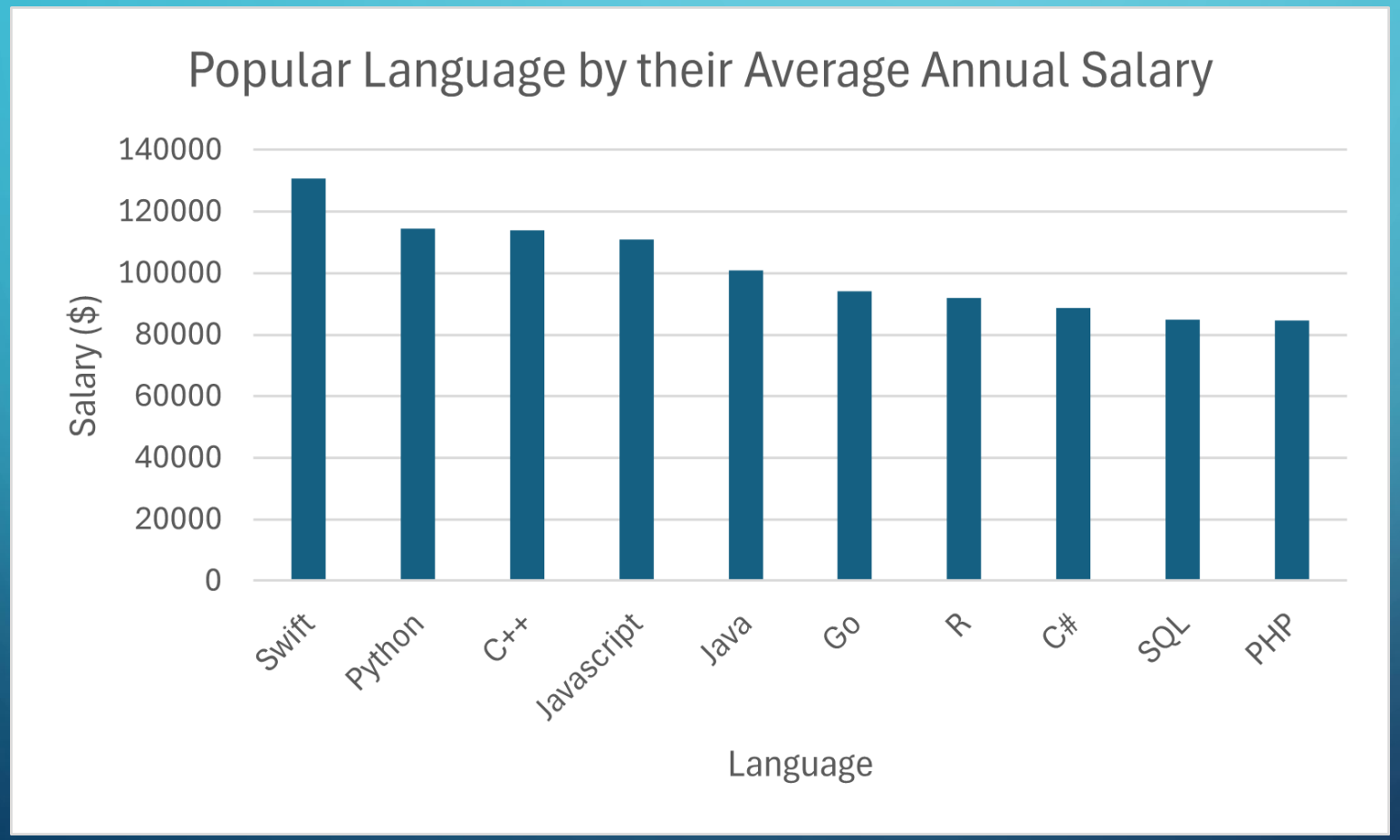

Fig. 6: Bar chart of Popular Languages

The figure demonstrates the bar chart of most popular languages by their annual average salary.

- Swift outperformed the list in most popular languages by the average salary of around $130000.
- Python and C++ comes in second and third in the list of most popular languages with almost equal amount of salary with $115000.
- Javascript can also be seen as a most famous language by in the list, however PHP comes at last in the list with around $90000.

DASHBOARD: CURRENT TECHNOLOGY USAGE

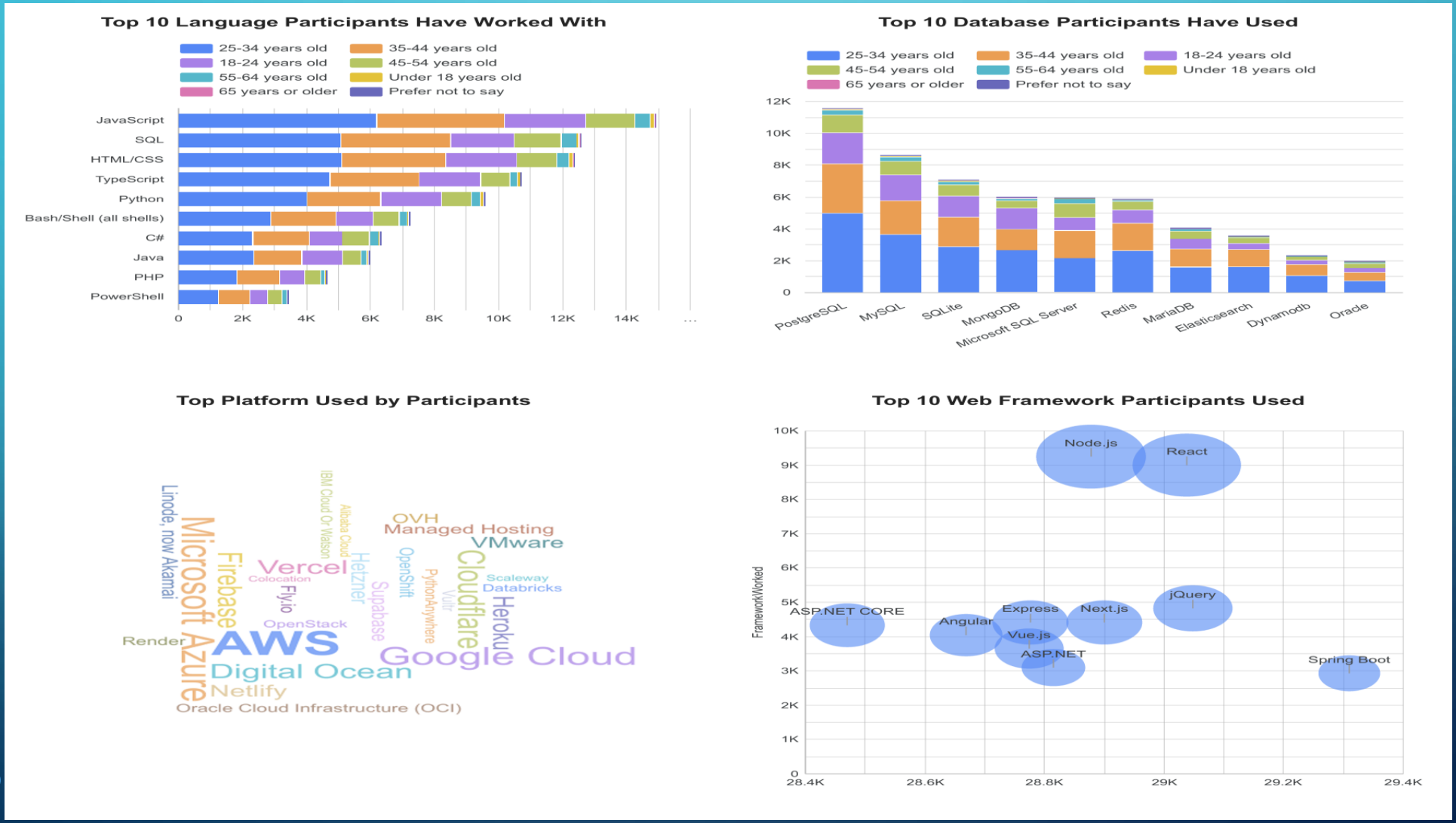

Fig. 7: Dashboard for Current Technology Usage, Looker Studio

DASHBOARD: FUTURE TECHNOLOGY TREND

Fig. 8: Dashboard for Future Technology Trend, Looker Studio

DASHBOARD: DEMOGRAPHICS

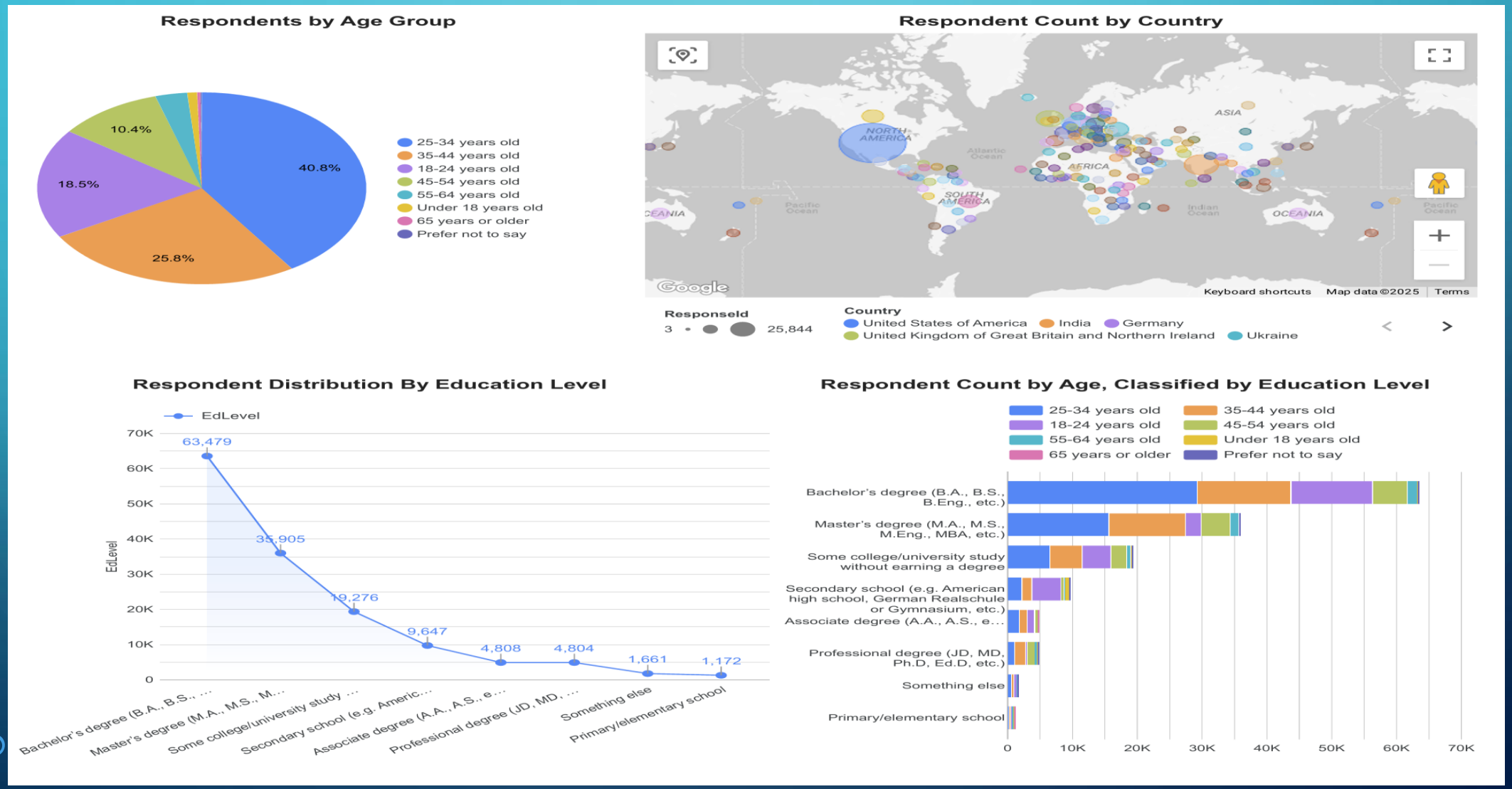

Fig. 9: Dashboard for Demographics, Looker Studio

DISCUSSION

- JavaScript, SQL and Python are foundational and consistently popular for both current usage and future desire. However there is a clear trend towards modernising these stacks with TypeScript, PostgreSQL, Go and Rush gaining significant grip for future adoption.
- AWS, Google Cloud and Microsoft Azure are the undisputed leaders in both current usage and future demand for cloud platforms, results in massive rely on cloud infrastructure.
- React and Node.js are most widespread and desired web framework in modern web application development, where Next.js is also showing strong future interest.
- Major workforce from around the world are in between 25-44 age groups, highly educated holding Bachelor's or Master's degree and are located in United States, India and Germany.
- Language like R, Java, Python and SQL has the highest number of job posting in the market where Swift, Python and C++ has become the most popular language in the developer market.

OVERALL FINDINGS AND IMPLICATIONS

- The analysis in the report found a dual trend in the developer ecosystem. A strong foundation in JavaScript, SQL and Python can be seen, yet there is a clear move towards modernising with Typescript, PostgreSQL, Go and Rust for future adoption. Cloud platforms like AWS, Google Cloud and Azure are dominant in cloud market, while React and Node.js rule web development, with Next.js gaining significant popularity.
- The significant workforce are young, highly educated, full-time employed, and globally distributed across US, India and Germany.
- Programming Language like R and Java see high job posting volume, Swift, Python and C++ command the popularity with highest salaries, indicating a market that values both widespread utility and specialised, high-value expertise.

CONCLUSION

- Languages like JavaScript, SQL and Python remain critical, but future growth points to modern alternatives like TypeScript, PostgreSQL, Go and Rust.
- Amazon Web Service (AWS), Google Cloud and Microsoft Azure remain prevalent.
- Workforce is predominantly young, highly educated and globally distributed with compensation strongly linked to experience.
- Finally, the industry demand continuous skill adaption, particularly in web, cloud and data technologies to align with evolving market needs and maximise career potential.

APPENDIX

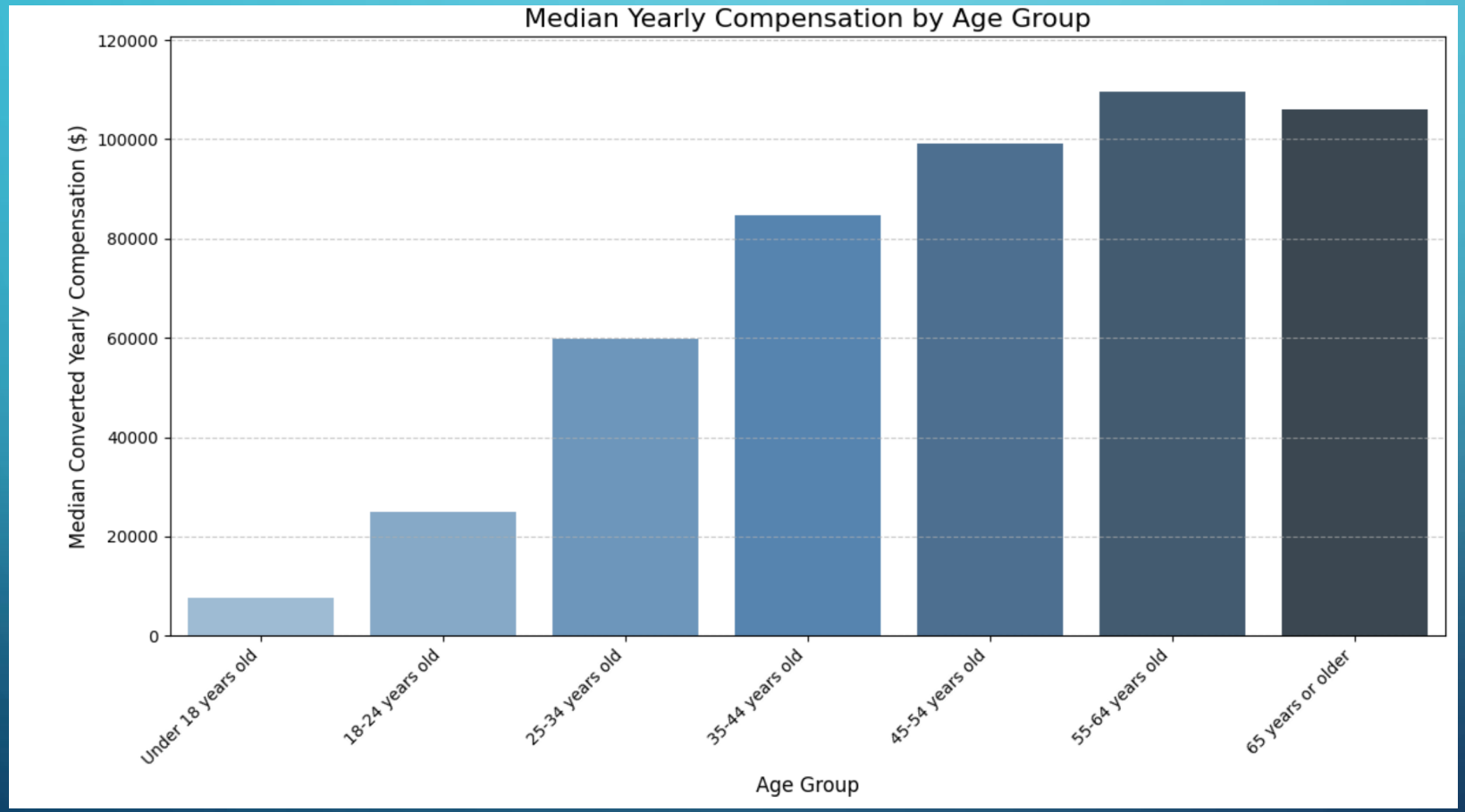

Fig. 10: Bar Graph of Yearly Compensation by age group

The bar plot illustrates the median yearly compensation by different age groups.

- Under 18 years old age groups earns the lowest amount of compensation and hold at the last at list of compensation by age group.
- The largest amount of compensation hold by the age group of 55-64 years old. This is positive relation between age and compensation as the work experience increases, increase the compensation as well.

APPENDIX cont.

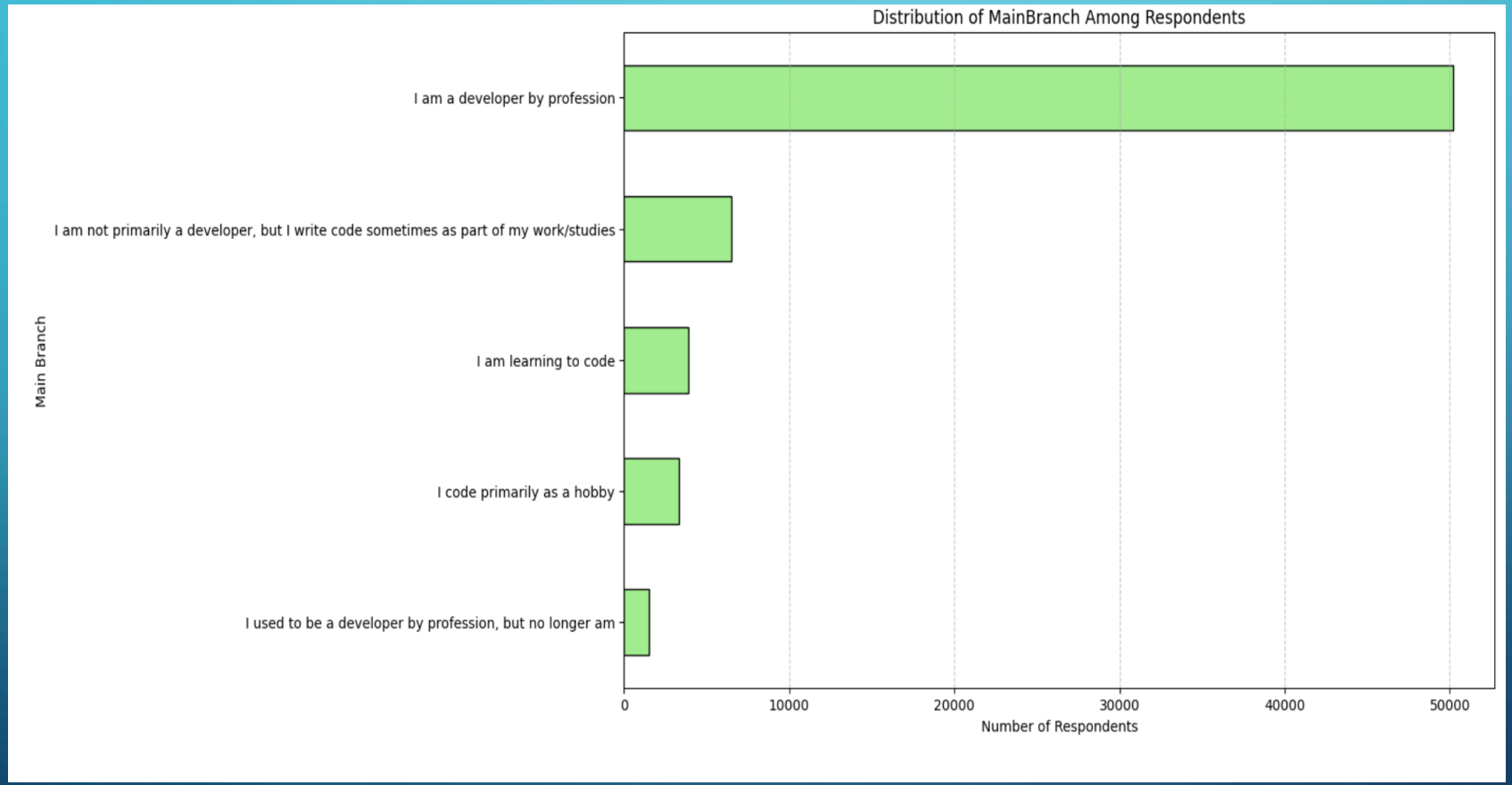

Fig. 11: Bar Graph of MainBranch amoung respondents

- The horizontal bar graph illustrates the vast majority of survey respondents are professional developers, significantly outnumbering other categories.
- While a strong range of professionals exists, there's also a notable presence of individuals who code as part of their work/studies, those learning to code, and hobbyists.

APPENDIX cont.

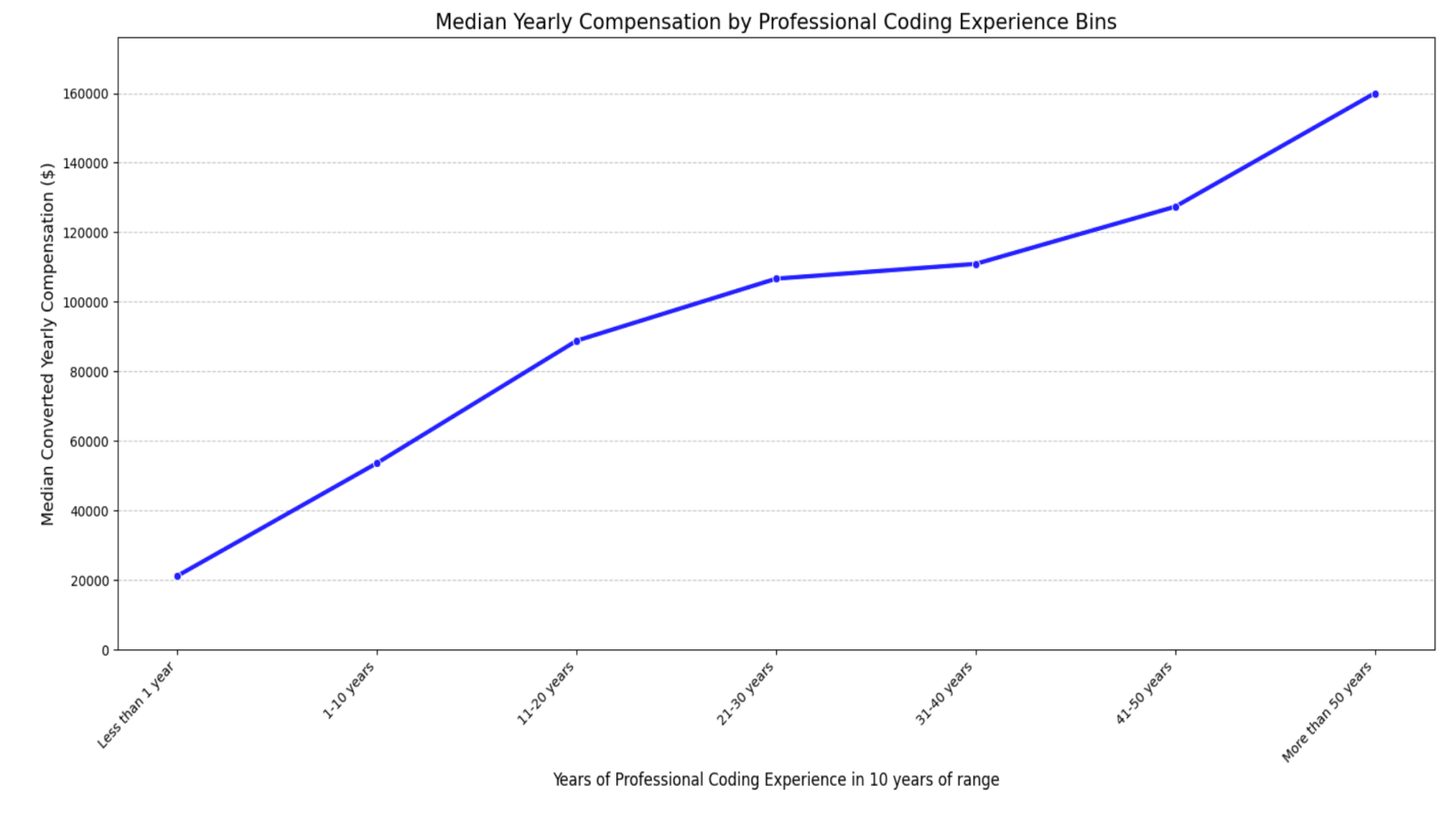

Fig. 12: Line Chart of Yearly Compensation by Experience

- This line chart demonstrates a clear and consistent positive correlation between professional experience and median yearly compensation.
- Experience data is binned into category for better visualization of the chart.
- Earnings generally increase significantly with experience, with the highest median compensation observed for those with "More than 50 years" of experience.

LINKS

GitHub link of the report: https://github.com/upbimal/IBM_Data_Analyst_Capstone_Project/
Google Looker Studio: https://lookerstudio.google.com/reporting/431551e7-94f0-4d6e-9ab8-d1f0847d4fe8