



# Data Analytics World from IBM

Wilson FLEURIMOND

2024-06-03

# OUTLINE

---



- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

---



- **Foundation of Data Analytics:**
  - Introduction to data analytics and its significance in modern business.
  - Overview of the data analytics lifecycle.
- **Data Management and Preprocessing:**
  - Techniques for data collection, data cleaning, and data integration.
  - Tools and best practices for data management.
- **Statistical Analysis and Visualization:**
  - Descriptive and inferential statistics.
  - Data visualization techniques using IBM Cognos Analytics and other visualization tools.
- **Advanced Analytics and Machine Learning:**
  - Introduction to machine learning concepts and algorithms.
  - Practical application of predictive analytics using Azure platform and IBM Watson.
- **Real-World Applications and Case Studies:**
  - Case studies demonstrating successful data analytics projects.
  - Hands-on projects to apply learned skills in real-world scenarios.

# INTRODUCTION

---



The IBM Data Analytics Training Program is designed to equip professionals with the necessary skills and knowledge to excel in the rapidly growing field of data analytics.

This comprehensive program covers a wide range of topics, from fundamental concepts to advanced techniques, ensuring that participants are well-prepared to tackle real-world data challenges.

The training includes hands-on experience with IBM's cutting-edge tools and technologies, fostering a deep understanding of data analytics processes and methodologies.

# METHODOLOGY



- **Point 1: Data Collection**

We sourced the data from various reputable online platforms and industry reports, ensuring a comprehensive and up-to-date dataset for our analysis. The main data sources included:

Online surveys of developers and IT professionals

Public datasets from technology and programming language communities

Industry reports and market research studies

- **Point 2: Data Preprocessing**

To ensure the data's quality and consistency, we performed several preprocessing steps, including:

**Cleaning:** Removing duplicates, handling missing values, and correcting inconsistencies in the data.

**Normalization:** Standardizing the data formats and scales for accurate comparisons.

**Integration:** Combining data from different sources to create a unified dataset for analysis.

- **Point 3: Data Analysis**

We applied various statistical and analytical techniques to uncover trends and insights, such as:

**Descriptive Statistics:** Summarizing the data to understand the general trends and patterns.

**Inferential Statistics:** Making inferences about the broader population based on sample data.

**Machine Learning Models:** Utilizing algorithms to predict future trends and identify patterns that are not immediately apparent.

# METHODOLOGY (bis)

- **Point 4: Data Visualization**

To effectively communicate our findings, we used advanced data visualization techniques, including:

**Bar Charts:** Displaying the popularity and usage trends of different programming languages and databases.

**Line Charts:** Showing the growth and decline trends over time.

**Pie Charts:** Illustrating the proportionate distribution of various factors, such as database usage during recession and non-recession periods.

**Bubble Charts:** Highlighting the impact of seasonality on automobile sales.

**Box Plots:** Identifying outliers and understanding the distribution of compensation data.

- **Sub Point 1: Tools Used**
  - We utilized Python libraries such as Matplotlib and Plotly for creating interactive and static visualizations
  - IBM Cognos Analytics was employed for creating dashboards and reports.
- **Sub Point 2: Visualization Techniques**
  - Applied customized filters to focus on the top trends and insights.
  - Used color-coding and labels for clarity and better understanding.

- **Point 5: Validation**

To ensure the accuracy and reliability of our findings, we employed several validation techniques:

**Cross-Validation:** Using different subsets of the data to validate the consistency of our machine learning models.

**Peer Review:** Engaging experts in the field to review our methodology and findings.

**Sensitivity Analysis:** Testing the robustness of our results by varying key assumptions and parameters.

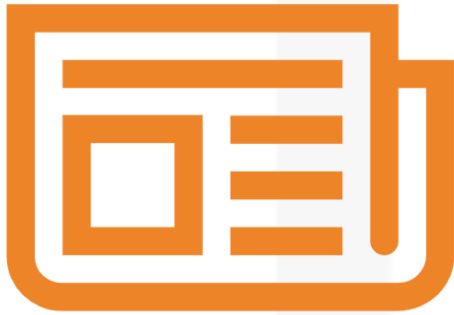
- **Point 6: Reporting**

We compiled our findings and visualizations into comprehensive reports and presentations to effectively communicate our insights to stakeholders:

**Executive Summary:** Highlighting key findings and implications.

**Detailed Analysis:** Providing an in-depth look at the data and the methodologies used.

**Recommendations:** Offering actionable insights based on our analysis to guide strategic decision-making



# RESULTS

---

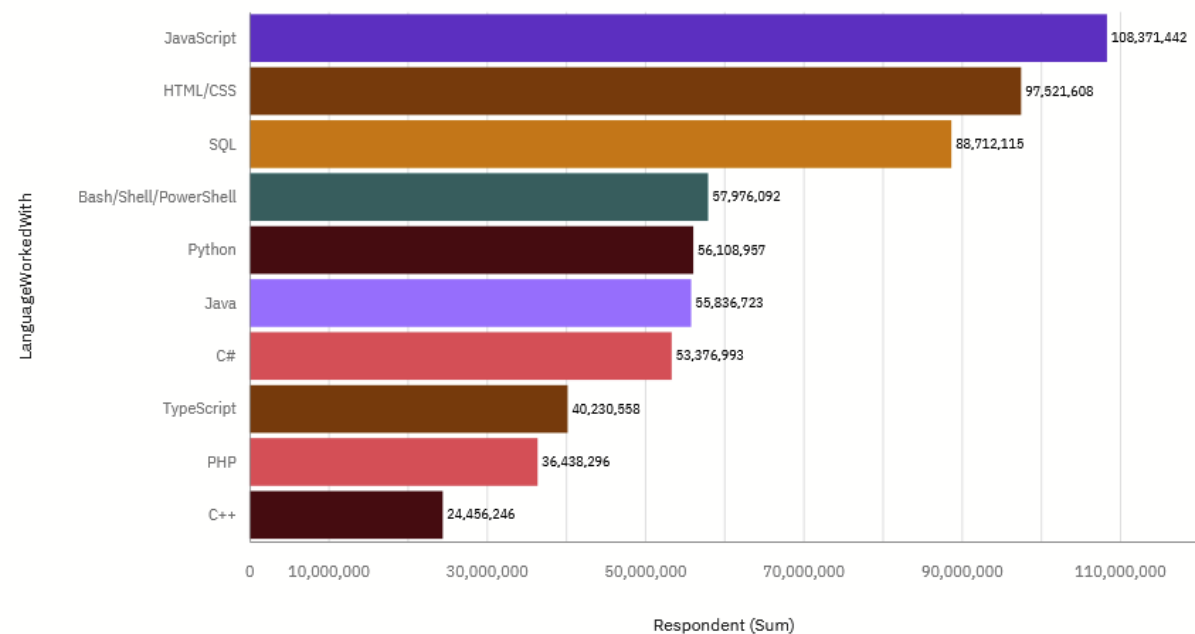
The IBM Data Analytics Training Program offers a robust pathway for professionals to develop and hone their data analytics skills.

By leveraging IBM's advanced tools and methodologies, participants will be well-equipped to transform data into actionable insights, driving business success and innovation

# PROGRAMMING LANGUAGE TRENDS

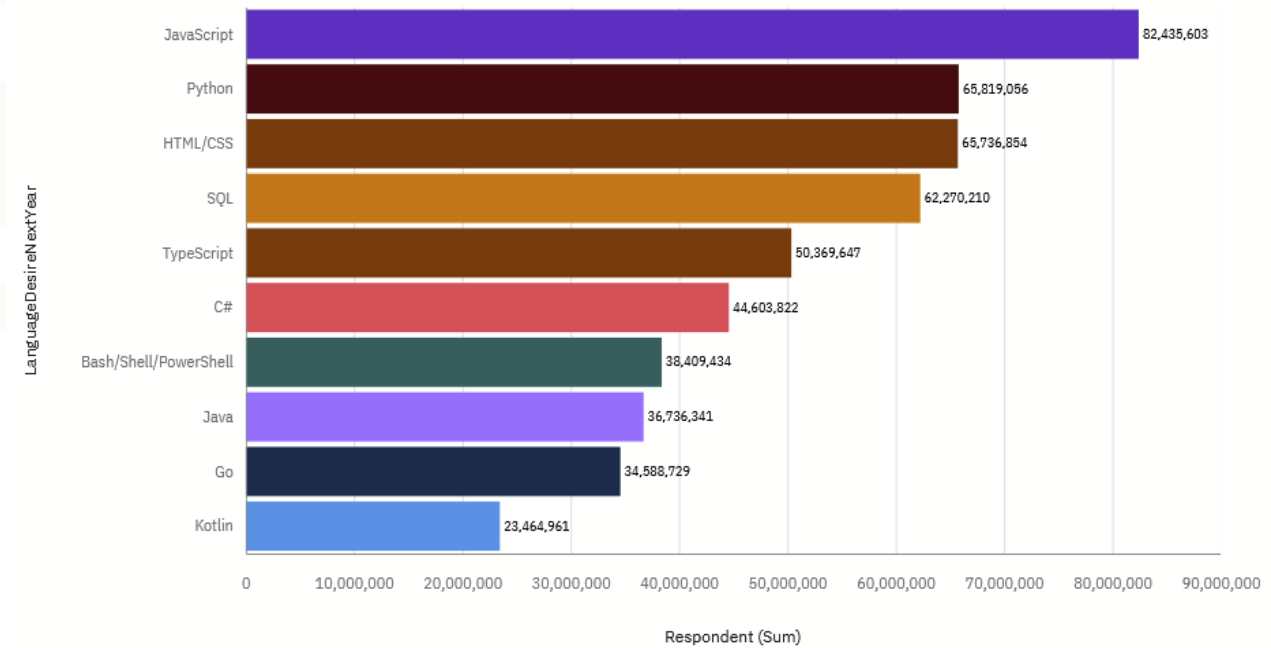
## Current Year

Top 10 LanguageWorkedWith



## Next Year

Top 10 LanguageDesireNextYear





# PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- **Finding 1:** Python is the most popular programming language among respondents, with a significant majority expressing their desire to learn or continue using it.
- **Finding 2:** JavaScript is highly favored for web development, showing strong preference among developers working on front-end and full-stack projects.
- **Finding 3:** Languages like Java and C# maintain a steady presence, particularly in enterprise environments, but newer languages such as Go and Rust are gaining traction due to their performance and concurrency features.

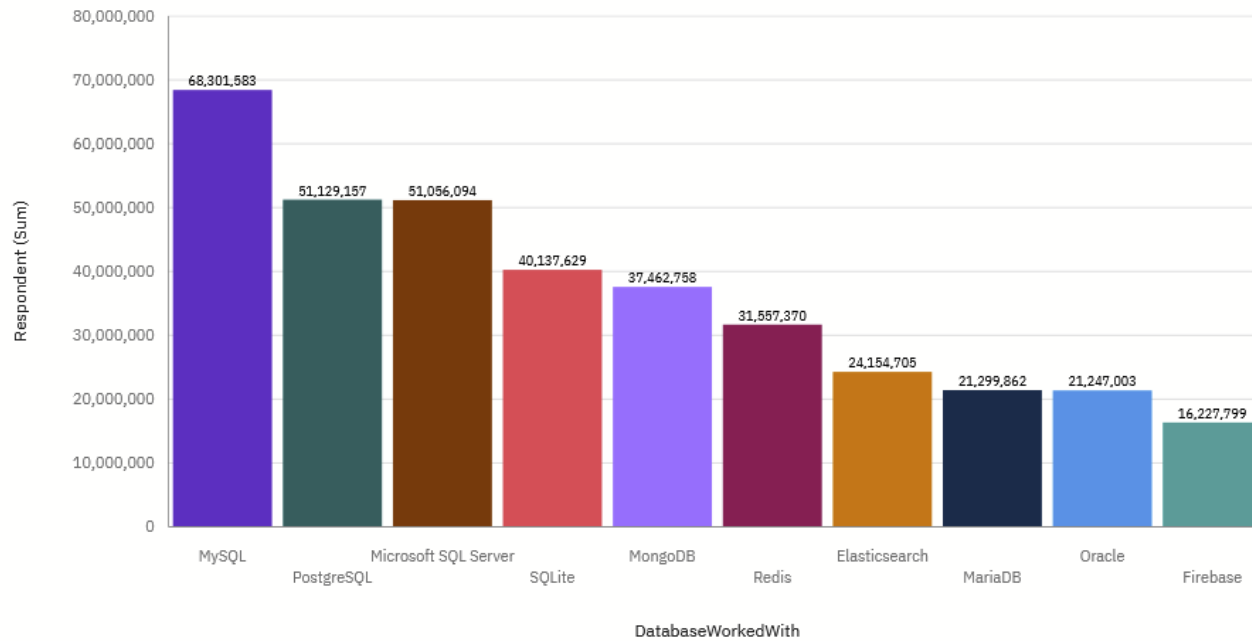
## Implications

- **Implication 1:** The rising popularity of Python indicates a strong trend towards data science, machine learning, and general-purpose scripting. Companies should consider investing in Python training and development resources to stay competitive and leverage these technologies.
- **Implication 2:** JavaScript's dominance in web development suggests that businesses focusing on web applications should prioritize JavaScript frameworks and libraries (e.g., React, Angular, Vue.js) in their tech stack to attract skilled developers and ensure efficient front-end development.
- **Implication 3:** The steady usage of Java and C# in enterprise solutions signifies that these languages will remain relevant for large-scale, legacy systems. However, the emergence of Go and Rust for performance-critical applications means organizations should also explore these languages to improve the efficiency and scalability of their systems, particularly for backend services and system programming.

# DATABASE TRENDS

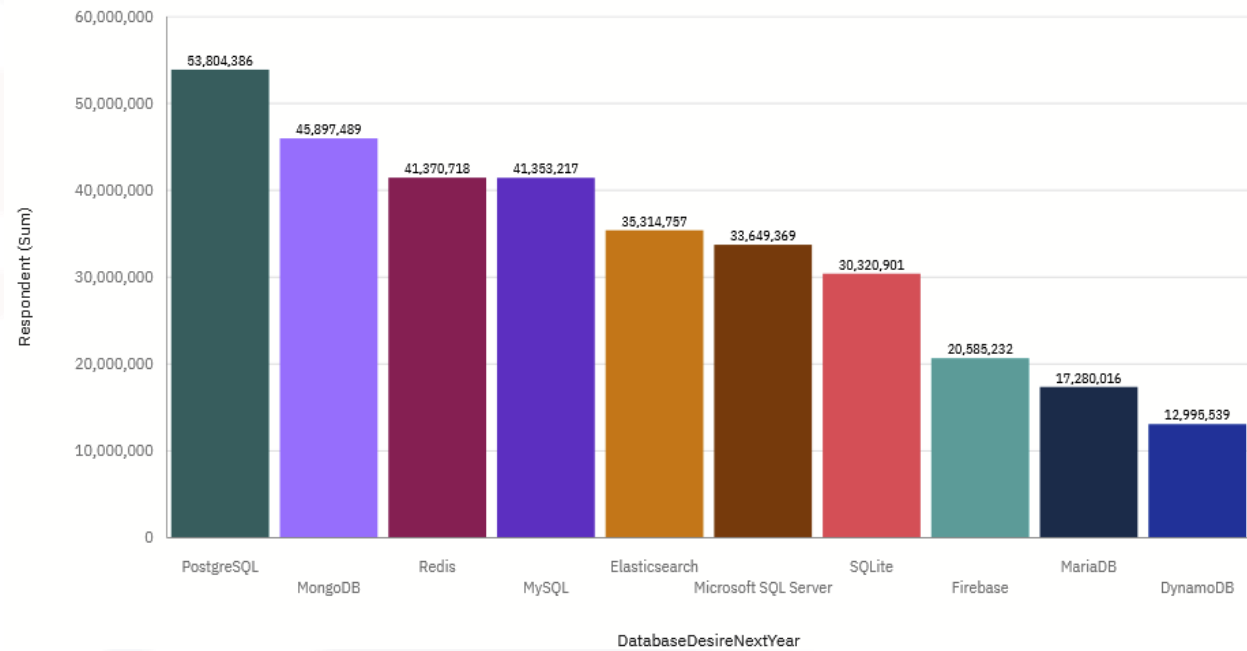
## Current Year

Top 10 DatabaseWorkedWith



## Next Year

Top 10 DatabaseDesireNextYear



# DATABASE TRENDS - FINDINGS & IMPLICATIONS

---

## Findings

- **Finding 1:** Relational databases such as MySQL, PostgreSQL, and SQL Server continue to be widely used, particularly in traditional business applications where structured data and ACID (Atomicity, Consistency, Isolation, Durability) properties are crucial.
- **Finding 2:** NoSQL databases like MongoDB and Cassandra are gaining popularity, especially in scenarios involving unstructured data, scalability requirements, and agile development practices.
- **Finding 3:** Cloud-based database services, including Amazon RDS, Google Cloud Spanner, and Azure SQL Database, are increasingly preferred due to their scalability, managed services, and reduced infrastructure overhead.

## Implications

- **Implication 1:** The continued use of relational databases indicates that businesses with established data infrastructures should maintain and optimize these systems to ensure reliable performance and data integrity. Investment in database administration and optimization tools remains important.
- **Implication 2:** The growing adoption of NoSQL databases suggests that organizations handling large volumes of unstructured data or requiring flexible schema designs should consider integrating NoSQL solutions into their data architecture. This will help in accommodating diverse data types and scaling horizontally as data grows.
- **Implication 3:** The shift towards cloud-based database services implies that companies should evaluate their database deployment strategies, considering cloud options to leverage benefits like automated maintenance, scalability, and cost efficiency. Migrating to or incorporating cloud databases can enhance business agility and reduce the burden of managing physical database servers.

# DASHBOARD

---



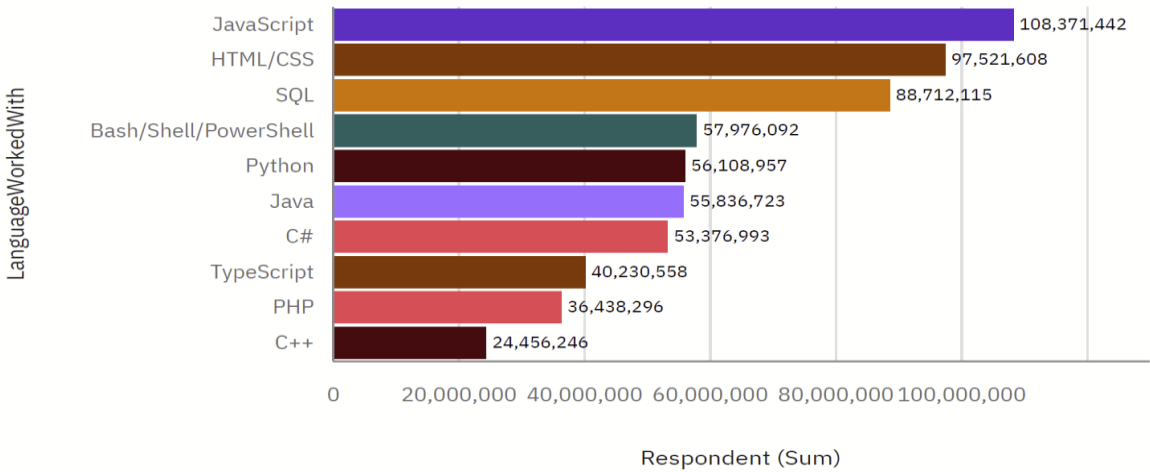
The GitHub link of my Cognos dashboard embedded :

<https://github.com/Letransparent/Technology-Usage-Visualization/blob/main/CognosDashboardEmbedded.ipynb>

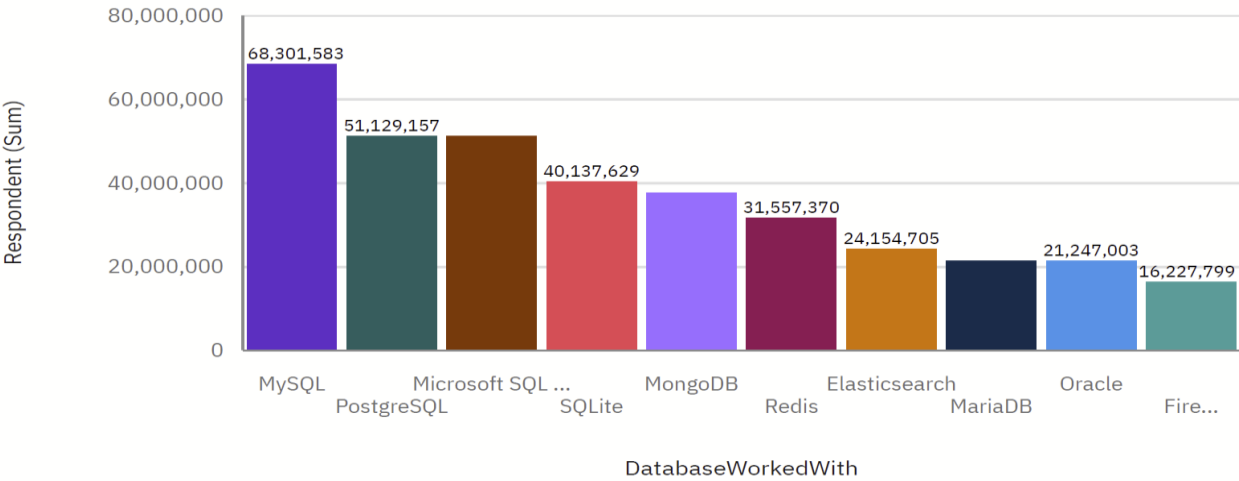
# DASHBOARD TAB 1

## Current Technology Usage

Top 10 LanguageWorkedWith



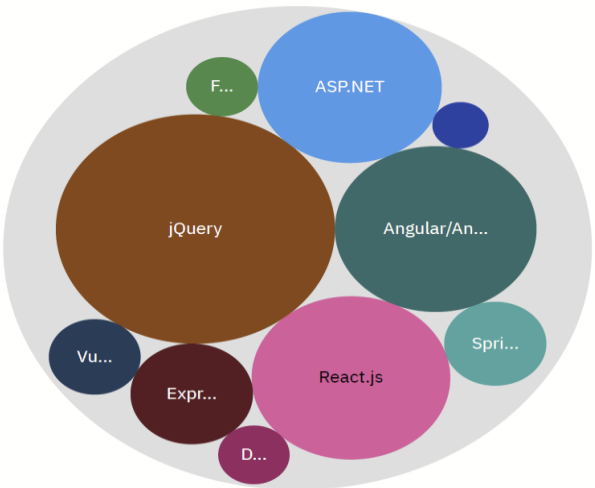
Top 10 DatabaseWorkedWith



Platform Worked With



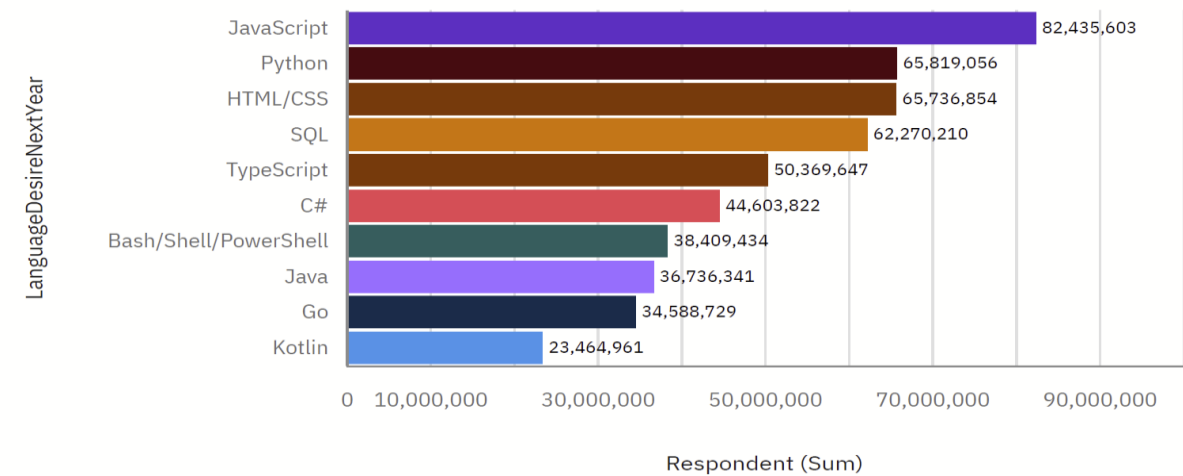
Top 10 WebFrameWorkedWith



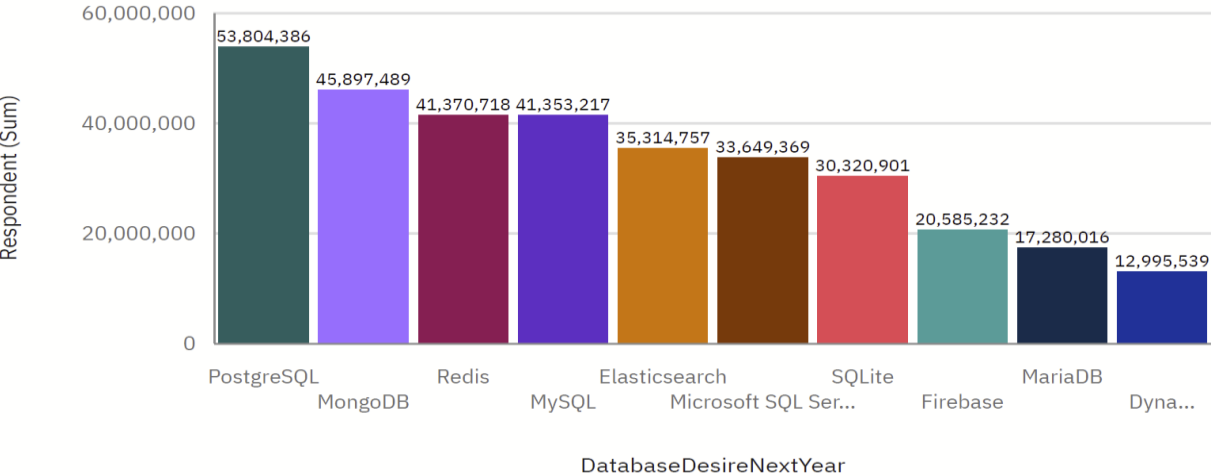
# DASHBOARD TAB 2

## Future Technology Trend

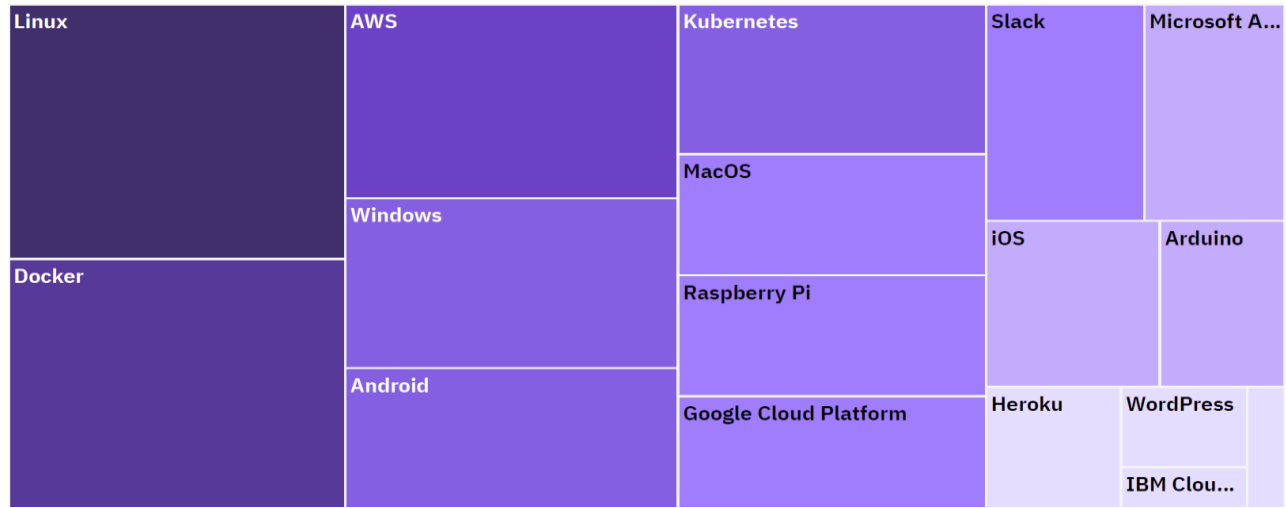
Top 10 LanguageDesireNextYear



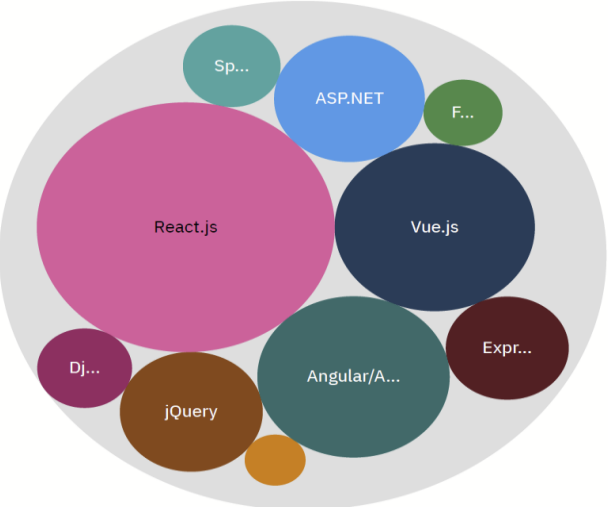
Top 10 DatabaseDesireNextYear



Platform Desire Next Year



Top 10 WebFrameDesireNextYear

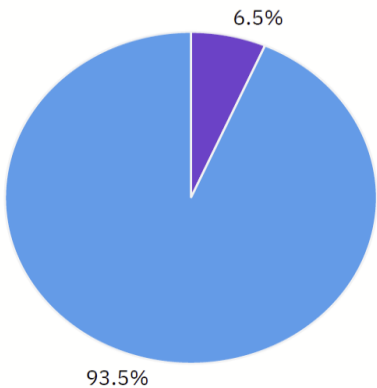


# DASHBOARD TAB 3

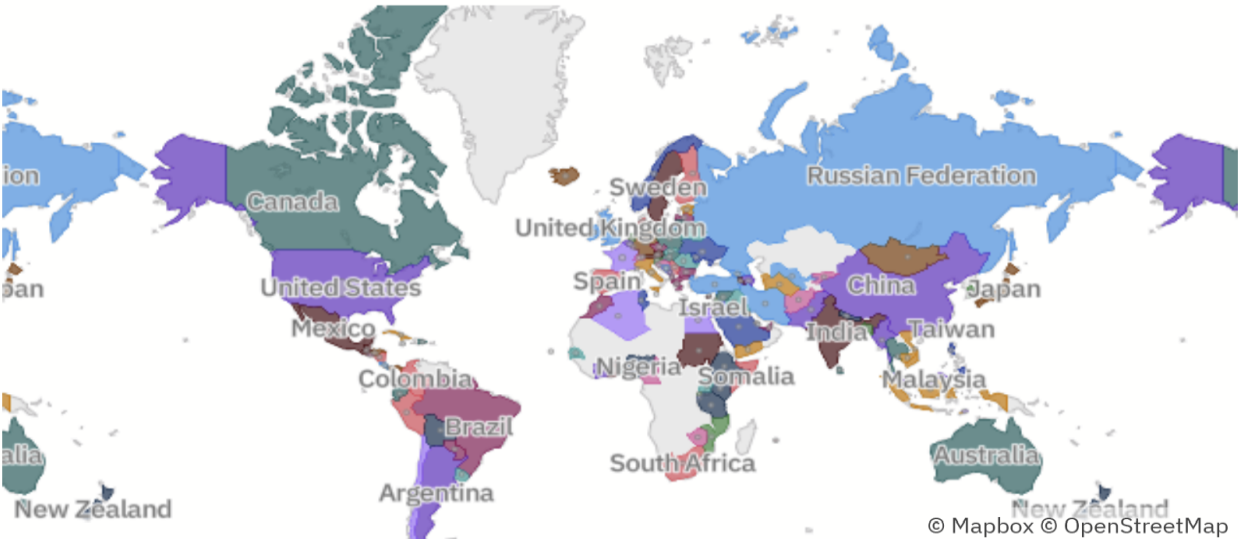
## Demographics

Respondent by Gender

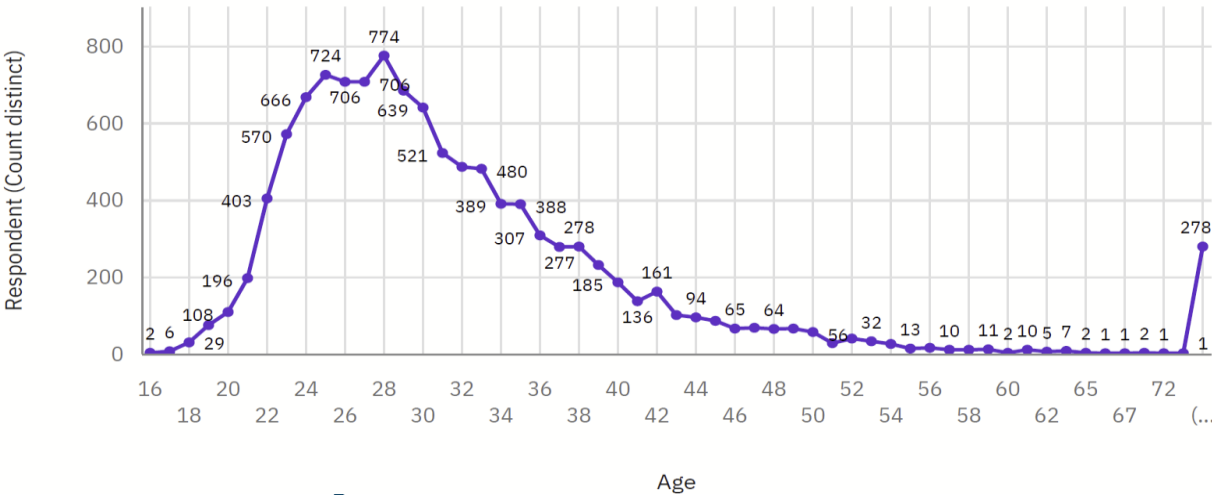
Gender  
● Woman ● Man



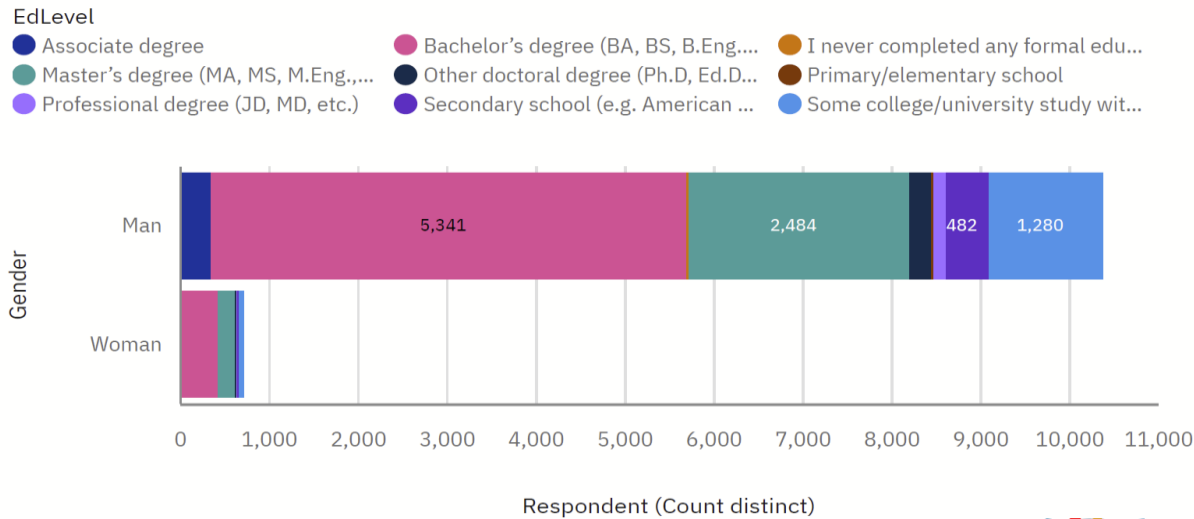
Country for Country regions and Respondent for Country points



Respondent by Age



Respondent Count by Gender, classified by Formal Education Level



# DISCUSSION

---

Our analysis of programming language and database trends has revealed several important findings that have significant implications for both developers and organizations.

- Organizations must be agile and forward-thinking in adopting new technologies. Keeping pace with advancements can provide a competitive edge and improve operational efficiency.
- For developers, continuous learning and skill development are crucial. Familiarity with multiple programming languages and database systems can enhance career prospects and enable them to tackle a broader range of projects.



# OVERALL FINDINGS & IMPLICATIONS

---

## Findings

- **Finding 1:** The demand for data analytics and data science skills continues to grow, with Python being the most sought-after programming language due to its versatility and extensive libraries for data manipulation and analysis.
- **Finding 2:** There is a significant trend towards the adoption of cloud-based solutions across various domains, including data storage, processing, and analytics. This shift is driven by the need for scalability, flexibility, and reduced infrastructure costs.
- **Finding 3:** Organizations are increasingly focusing on leveraging advanced analytics and machine learning to gain insights and drive decision-making. This includes the use of sophisticated algorithms and models to predict trends, understand customer behavior, and optimize operations.

## Implications

- **Implication 1:** Educational institutions and training providers should continue to emphasize Python in their curricula and offer specialized courses in data analytics, machine learning, and cloud computing to equip students with the skills needed in the current job market.
- **Implication 2:** Businesses should invest in cloud infrastructure and services to take advantage of their scalability and cost-efficiency. This includes migrating existing applications and data to the cloud, as well as developing new cloud-native solutions to enhance agility and innovation.
- **Implication 3:** Companies should prioritize building or expanding their data science teams and capabilities. This involves not only hiring skilled data scientists and analysts but also investing in tools and platforms that facilitate advanced analytics and machine learning. Additionally, organizations should foster a data-driven culture to ensure that insights derived from data are effectively integrated into strategic and operational decision-making processes.

# CONCLUSION

---

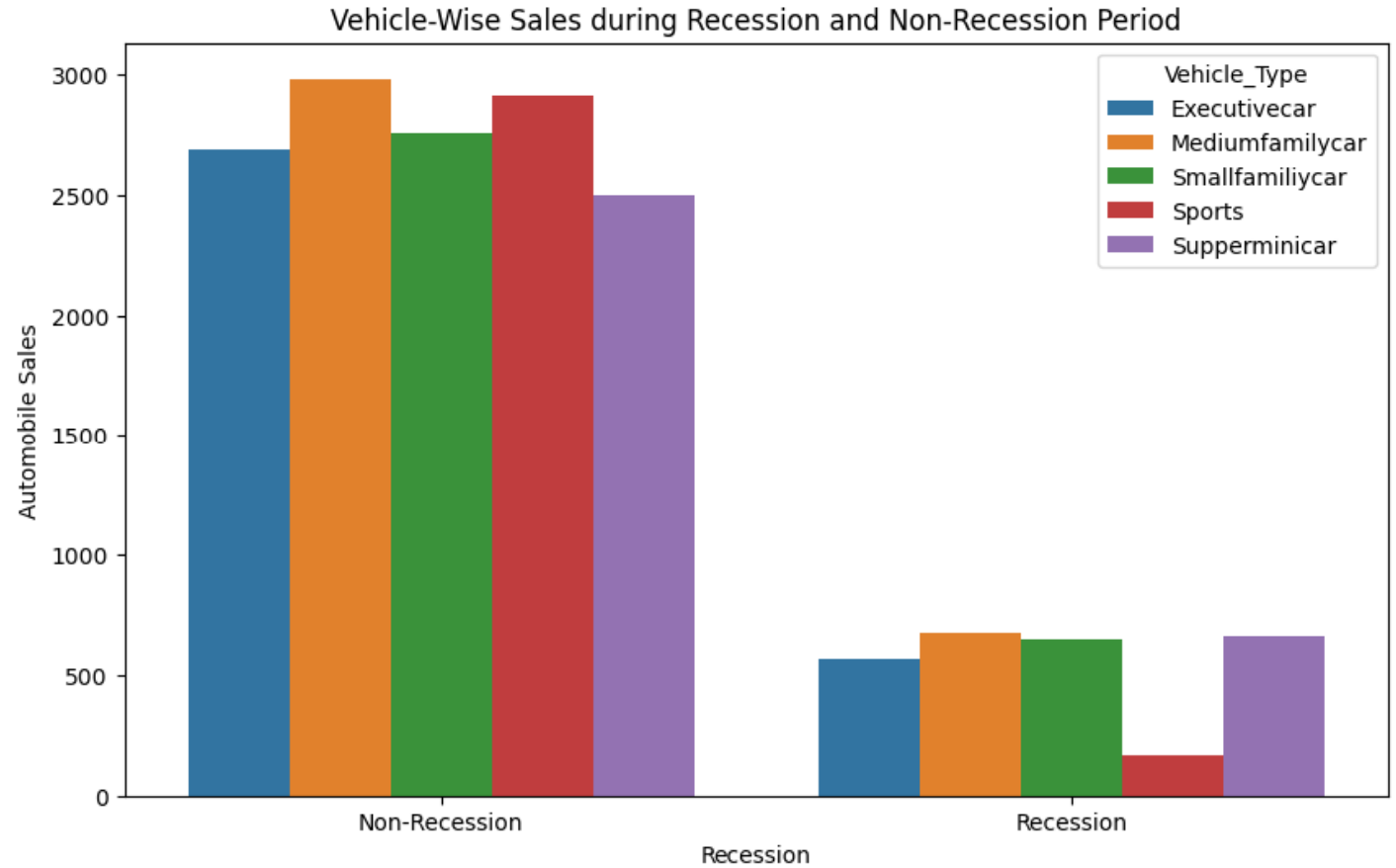


This analysis underscores the dynamic nature of the tech industry, highlighting the importance of adaptability and continuous learning.

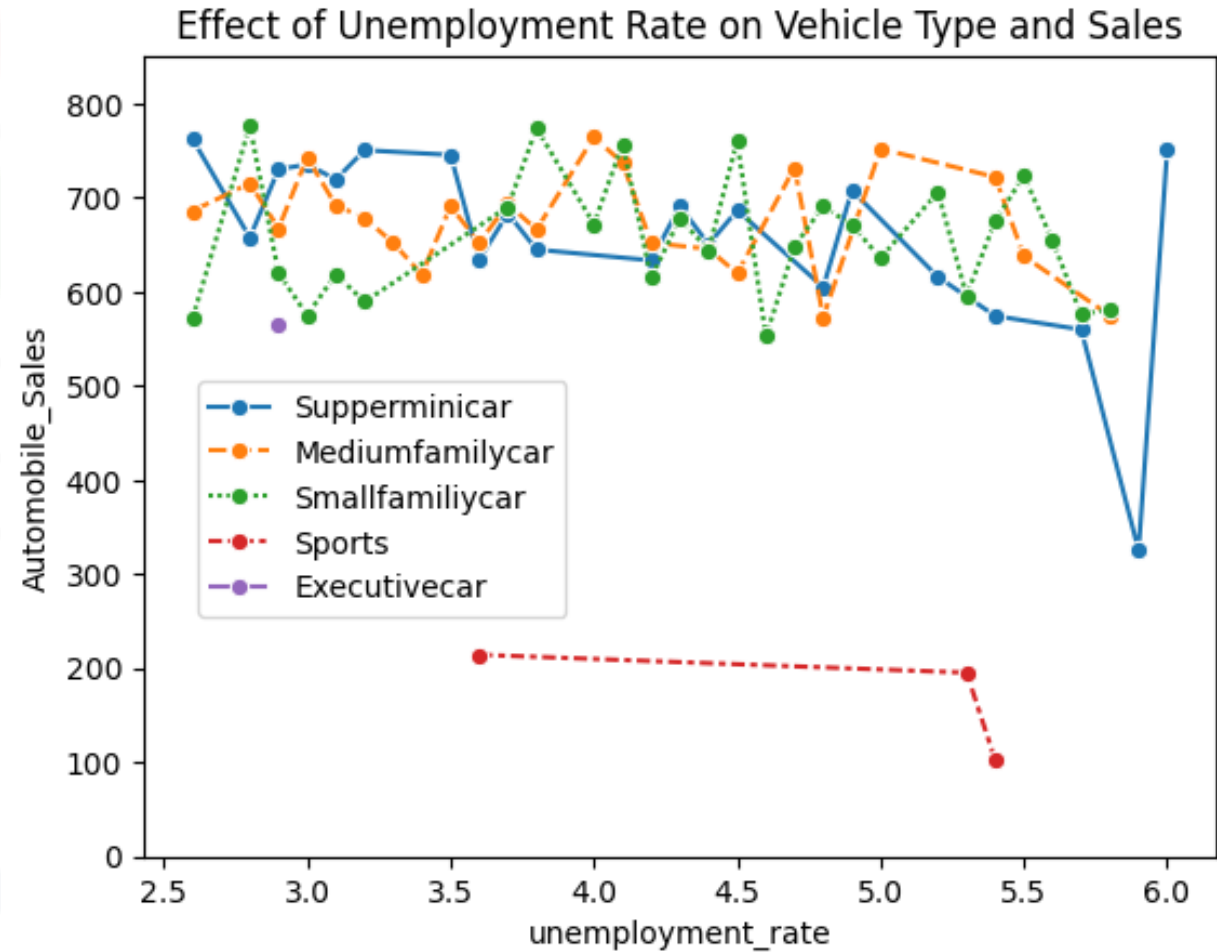
By understanding current trends and making informed decisions based on data-driven insights, organizations and developers can navigate the complexities of the tech landscape and capitalize on new opportunities.

As the industry evolves, staying proactive and embracing change will be crucial for sustained success and innovation.

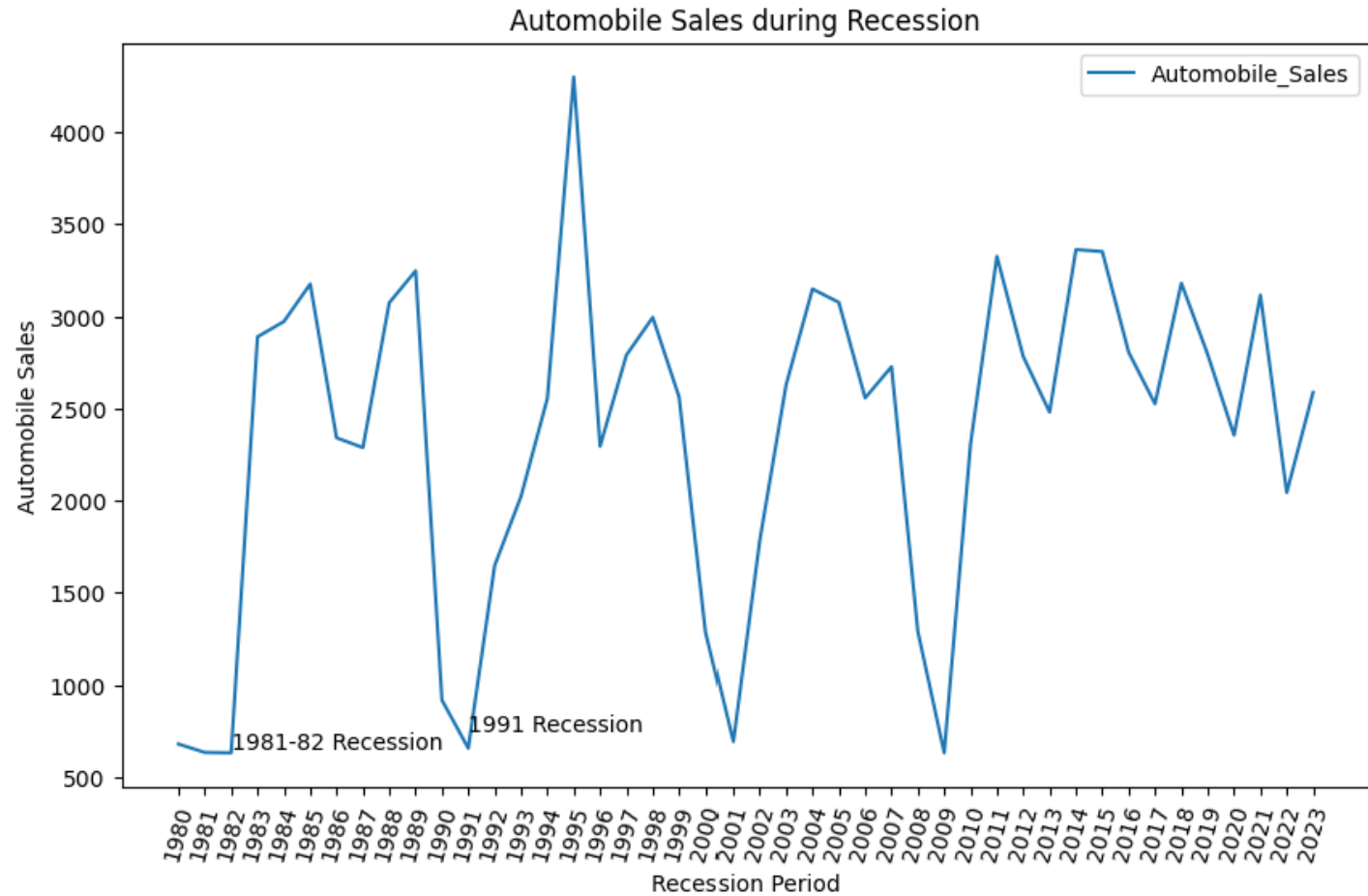
# APPENDIX



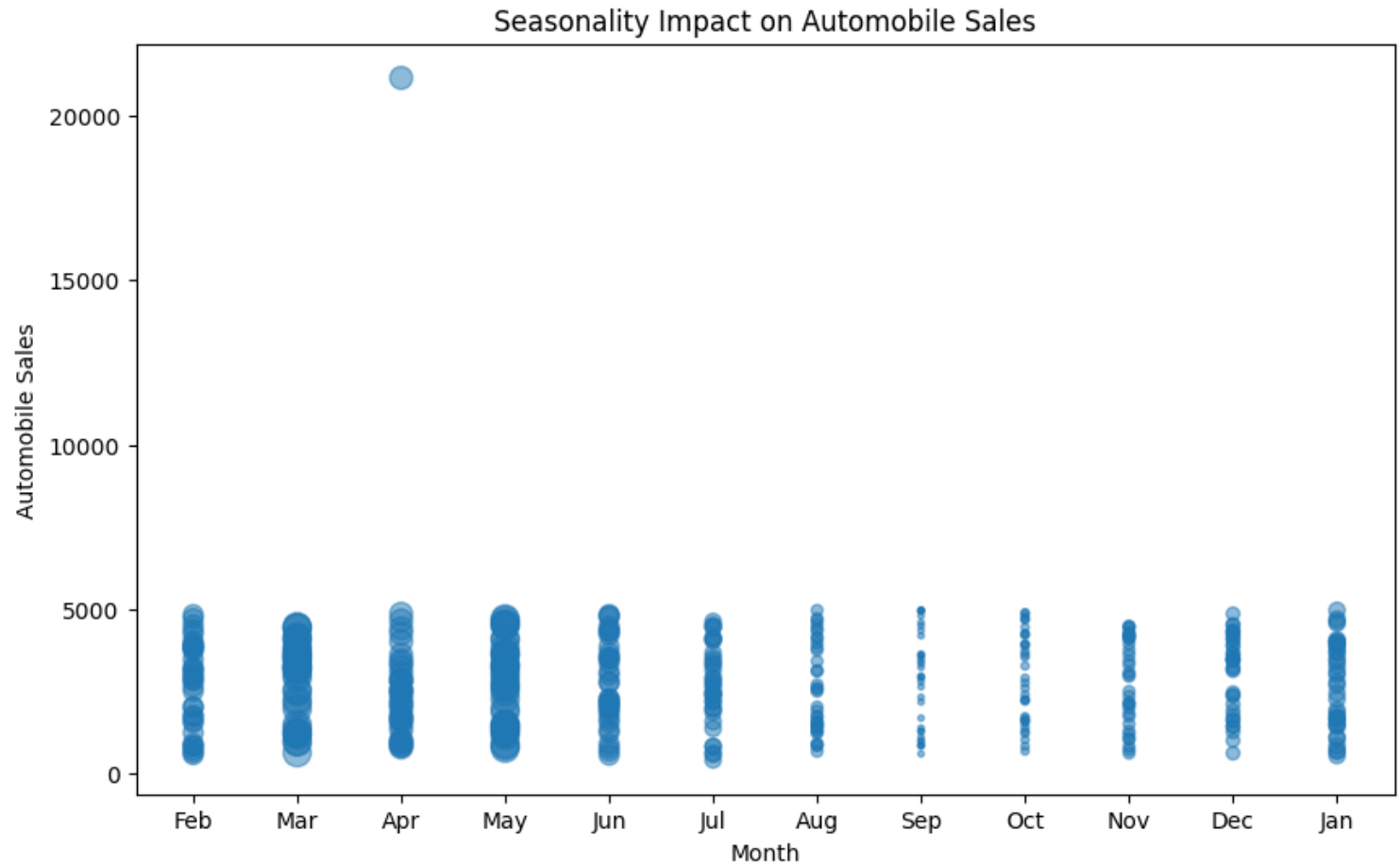
# APPENDIX



# APPENDIX

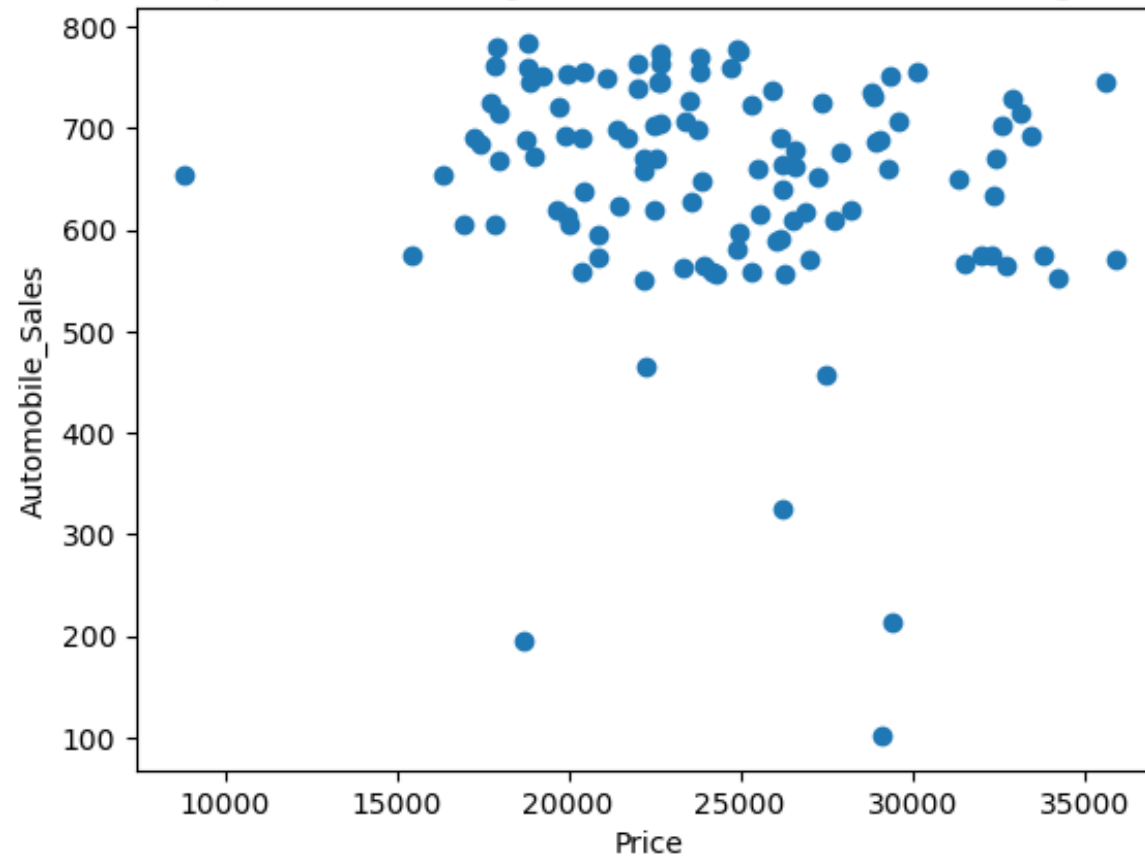


# APPENDIX



# APPENDIX

Relationship between Average Vehicle Price and Sales during Recessions

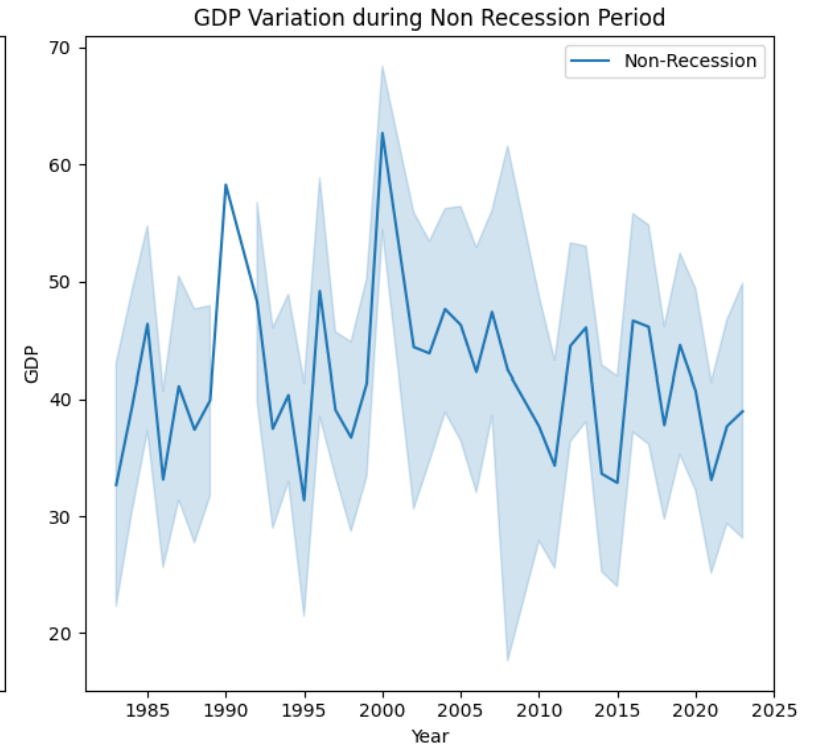
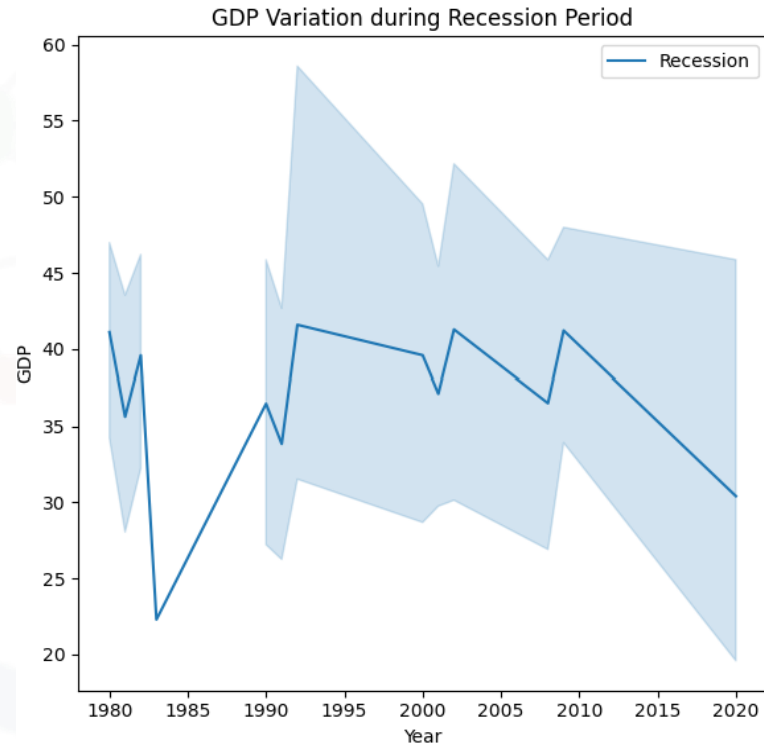


# APPENDIX

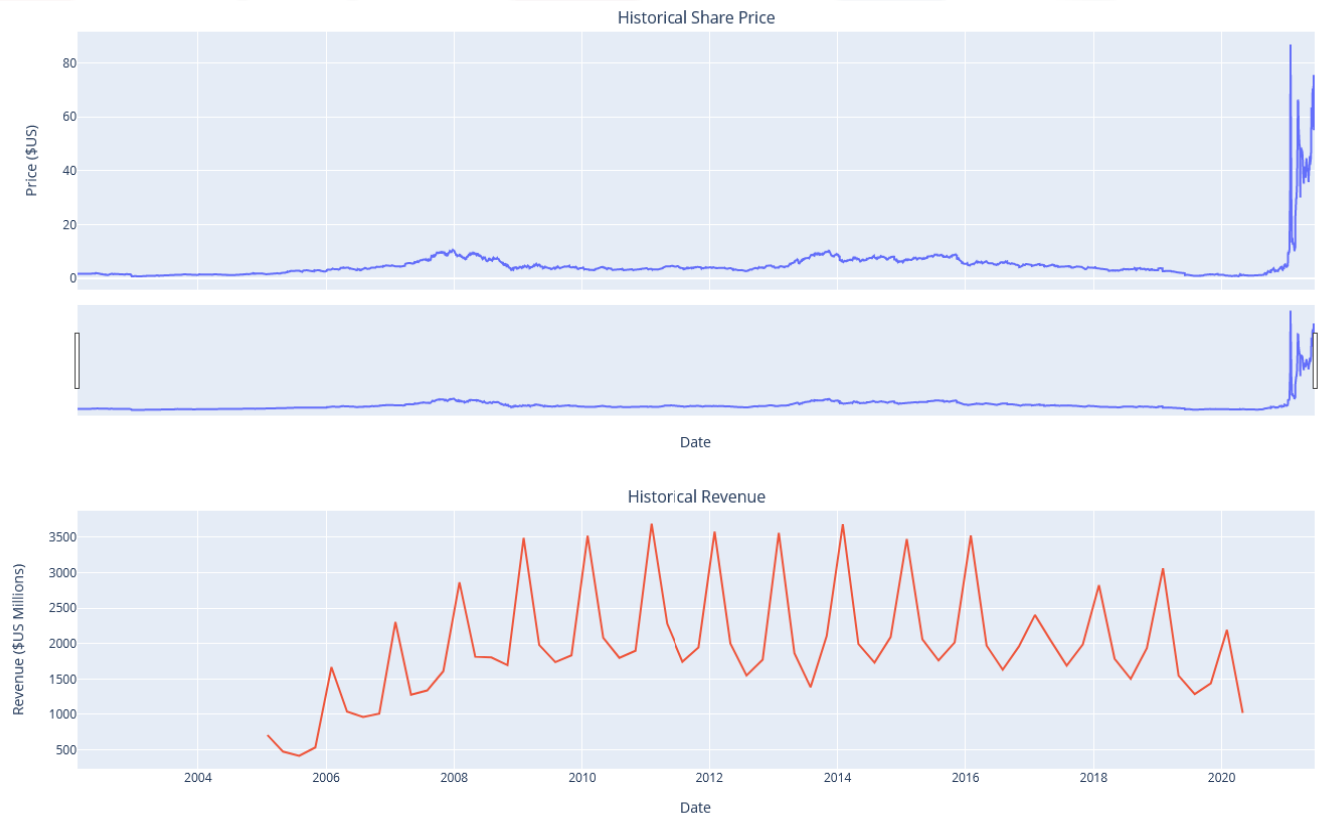




# APPENDIX



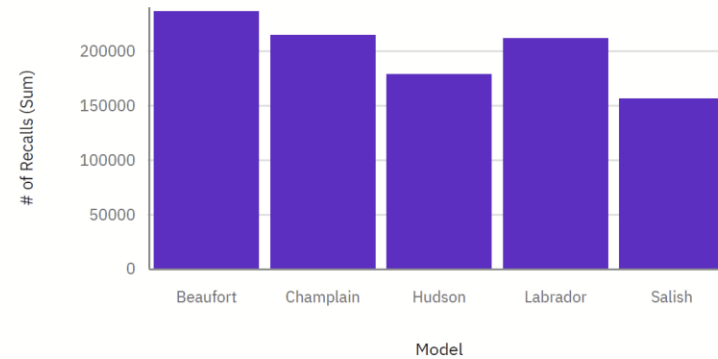
# APPENDIX



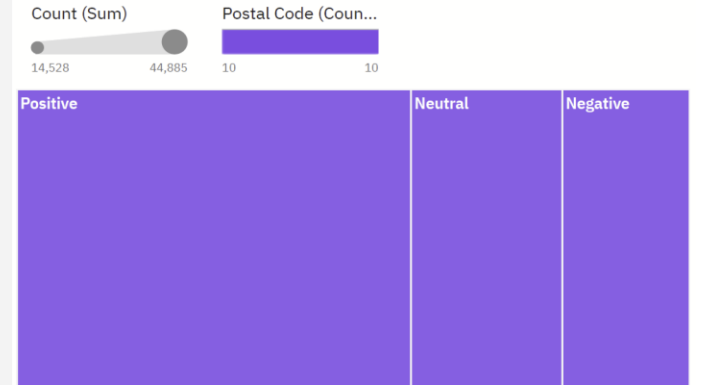
# APPENDIX

## Service

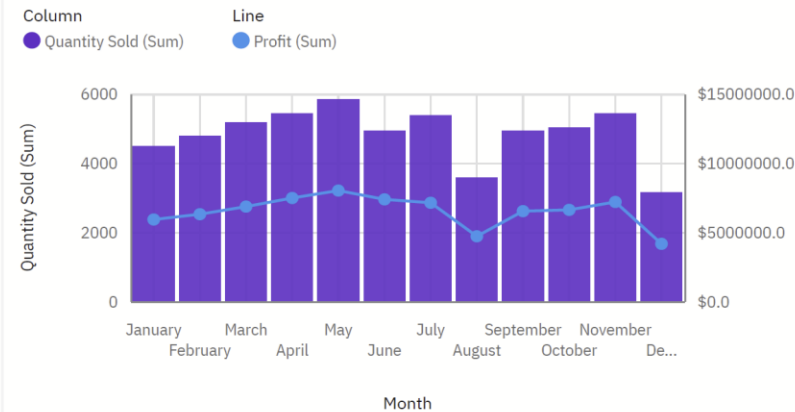
# of Recalls by Model



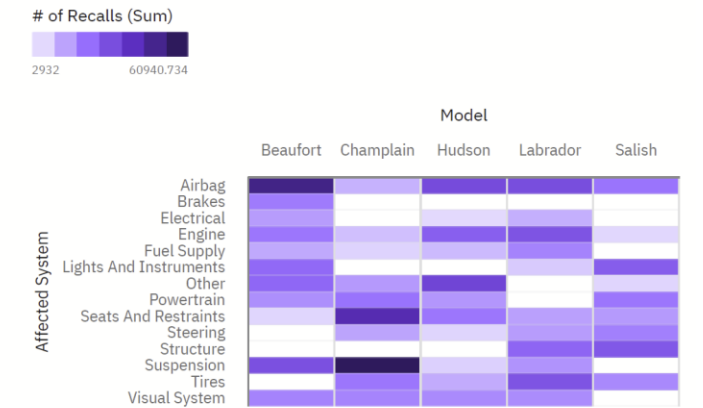
Sentiment hierarchy colored by Postal Code and sized by Count



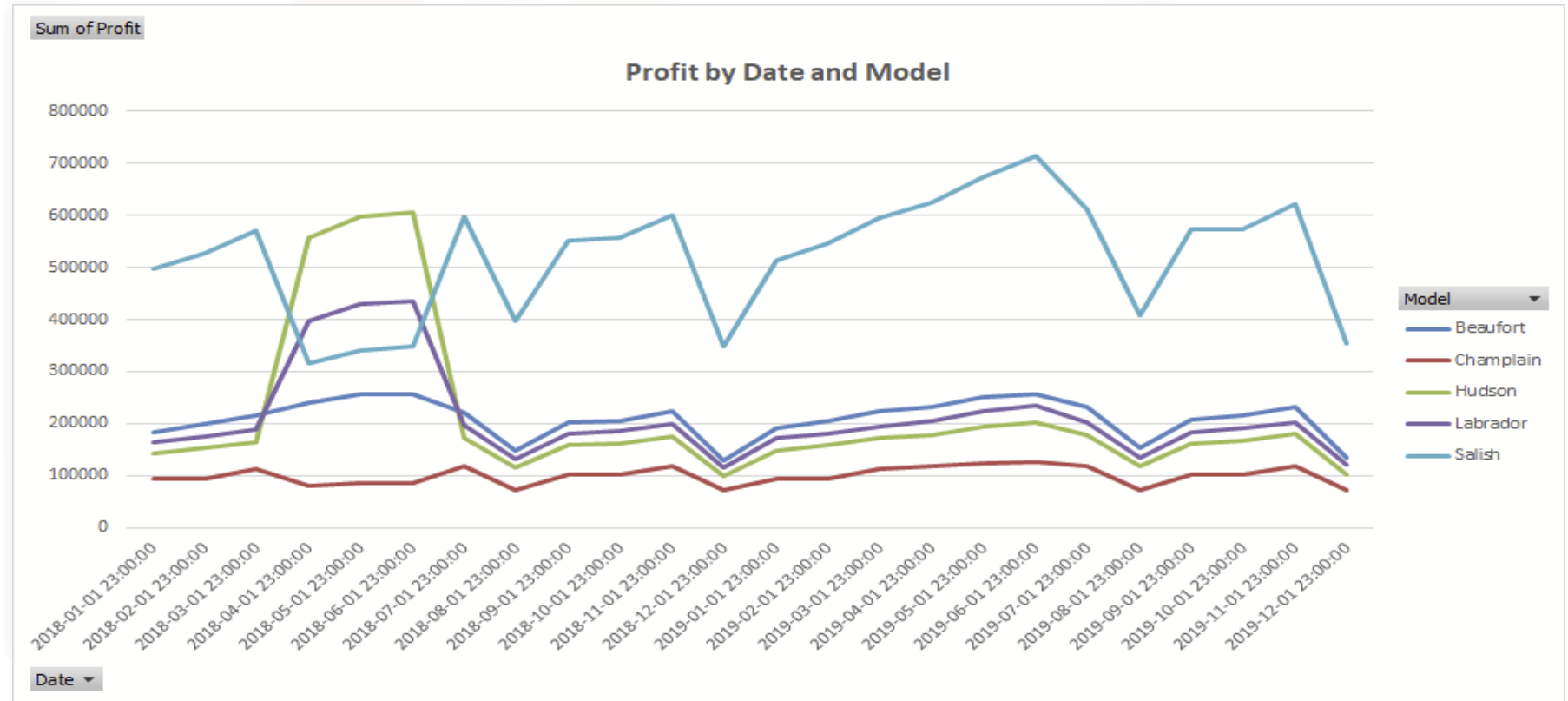
Profit and Quantity Sold by Month



# of Recalls by Affected System and Model

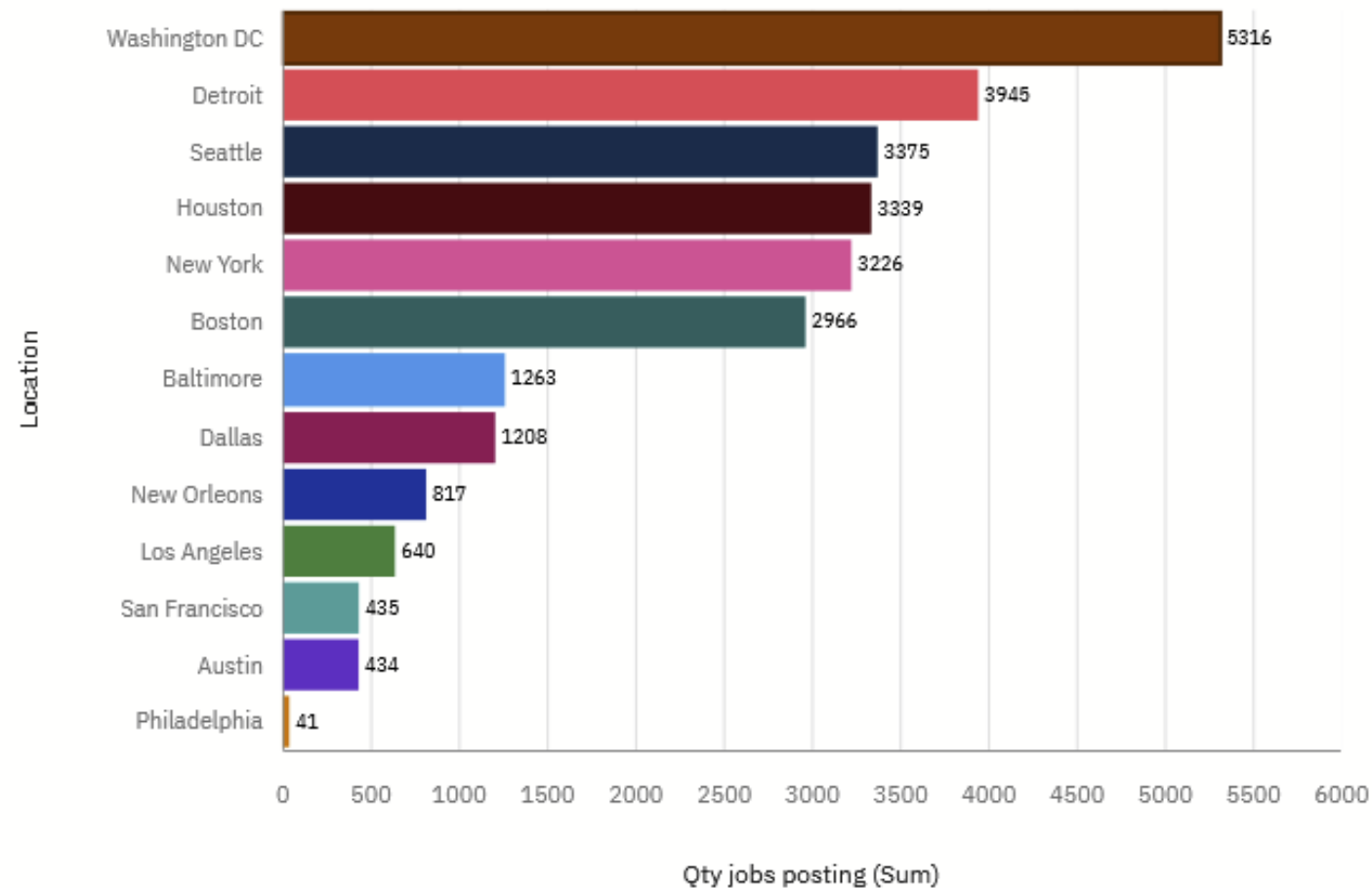


# APPENDIX



# JOB POSTINGS

Qty jobs posting by Location colored by Location



# POPULAR LANGUAGES

Average Annual Salary by Language colored by Language

