Languages, Databases & Demographics: Our Journey Forward

Andre Nel

August 28, 2024

OUTLINE

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

EXECUTIVE SUMMARY

- **Methodology:**
  - Survey with 20,000+ participants.
  - Data cleaned and anonymized.
- **Ranked Voting System:**
  - 100 votes allocated, max 10 per feature.
  - Captures nuanced user preferences.
- **Job Postings & Trends:**
  - Tracks industry demand for languages/databases.
  - Highlights geographic and industry-specific trends.

- **Survey Focus Shift:**
  - From languages/databases to feature preferences.

KAGGLE SURVEY METHODOLOGY

- **Overview:** Conducted by Kaggle to gather insights from global data science professionals.
- **Audience:** Data scientists, academics, and students across various experience levels.
- **Distribution:** Promoted via Kaggle, social media, and email to over 5 million users.
- **Duration:** Open for a few weeks with multiple reminders sent.
- **Data Collection:** Anonymous responses on demographics, tools, and practices.
- **Response Rate:** Over 20,000 participants from diverse regions.
- **Processing:** Data cleaned and anonymized; publicly availabl for research.
- **Reporting:** Results shared with the community in detailed reports.

INTRODUCTION PROPOSED SURVEY MODS

- **Kaggle's Previous Focus**
  - *Languages and Databases:* Survey centered on the tools and technologies used.
  - *Pros:*
    - Clear understanding of popular languages and databases.
    - Useful for benchmarking tool usage trends.
  - *Cons:*
    - Limited insight into user needs and feature preferences.
    - Overemphasis on technology rather than user experience.
- **Proposed Focus on Features**
  - *Shift to Features:* Survey now emphasizes desired features and user needs.
  - *Pros:*
    - Better alignment with user needs and expectations.
    - Facilitates the development of more user-centric tools.
  - *Cons:*
    - Potential for less emphasis on technological trends.
    - Requires careful interpretation of feature requests to avoid bias.

RANKED VOTING

RANKED VOTING SYSTEM WITH BUDGET

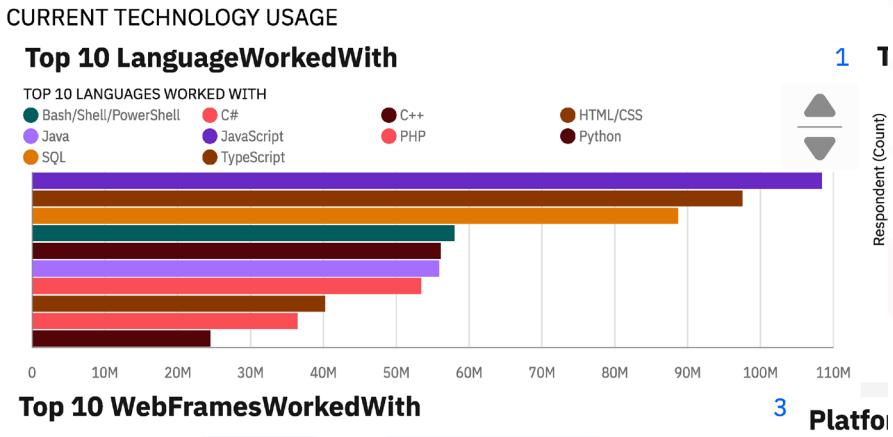
- **Voting Budget:**
  - Each voter is allocated 100 votes.
- **Distribution:**
  - Voters can allocate up to 10 votes per feature.
  - Must allocate all 100 votes across their selected features.
- **Purpose:**
  - Allows voters to express varying levels of preference.
  - Ensures comprehensive input by requiring full vote allocation.
- **Benefits:**
  - *Flexibility:* Voters can strongly support multiple features.
  - *Precision:* Captures nuanced preferences and priorities.

RESULTS

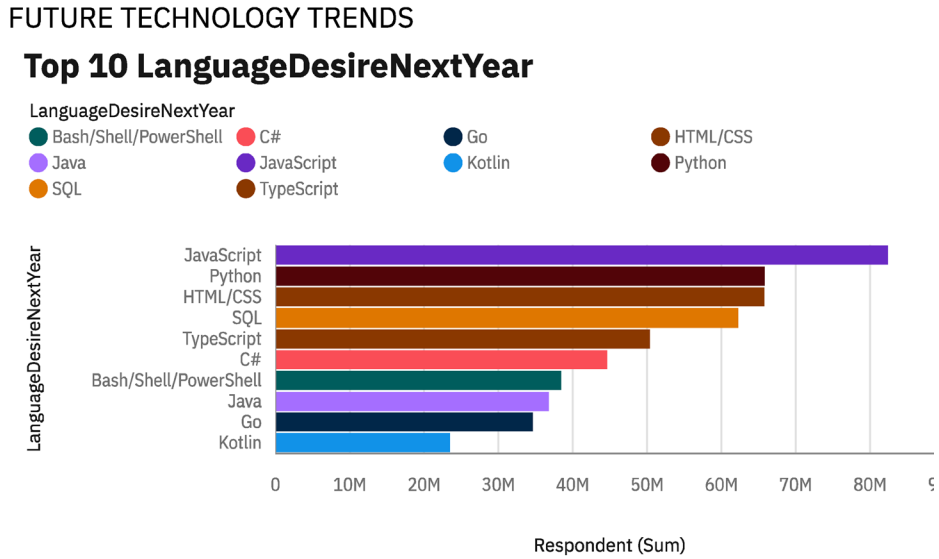
TBD

PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

Top 10 LanguageDesiredNext Year

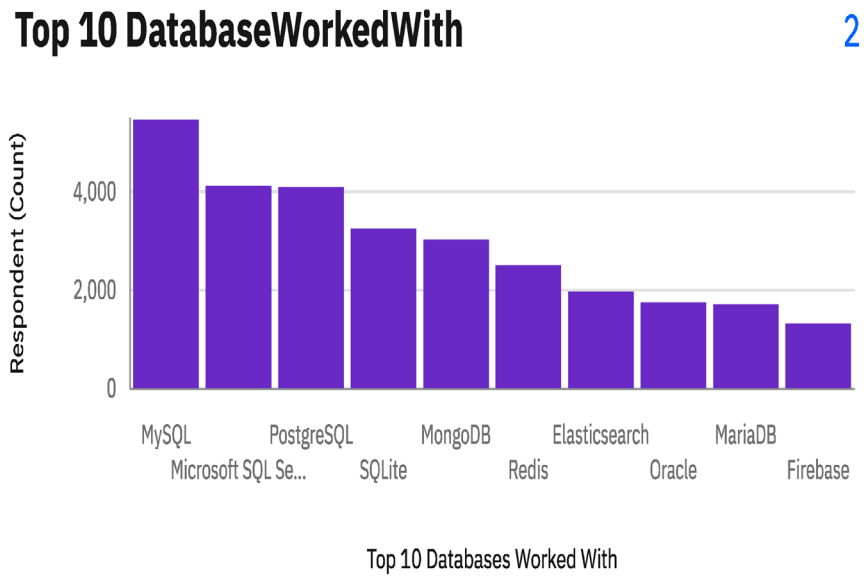
- C# Relative Rank Moved Up
- C Appears On List
- C++ Relative Rank Moved Up
- Go, Dart and Assembly Appear on List

Implications

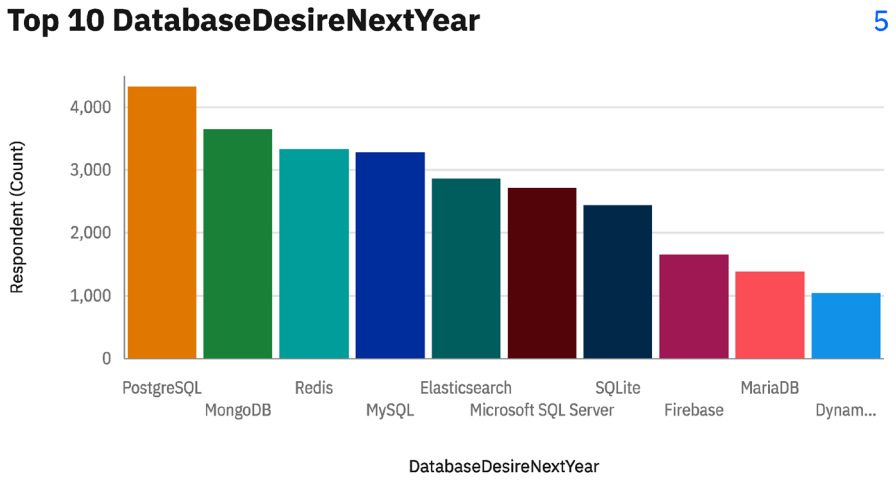
- Language Wants & Needs Change
- Need to Find Out What is Driving the Changes
- Need to Find Out If Changes are Warranted and Practical

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

Top 10 on Databases DesiredNextYear

- MySQL Moved From Top to Bottom of List
- ElasticSearh Moved From Bottom to Top
- Cassandra and Couchbase Appear on List

Implications

- Database Wants and Needs are Changing
- Need to Figure Out What is Driving Change
- Need to Determine is Change Practical

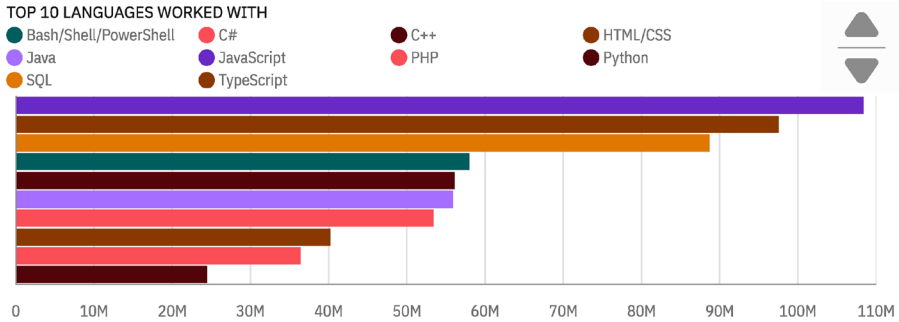
DASHBOARDS HERE

- https://github.com/amenel030350/Dashboard-Submission/commit/6cba11b4bdb414a10f5f2688fe306e8d5ae37ab4

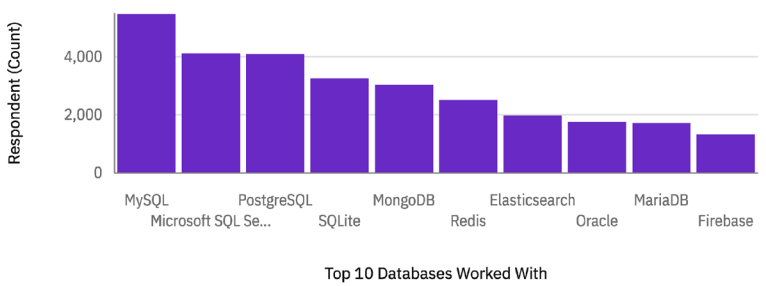
DASHBOARD TAB 1

CURRENT TECHNOLOGY USAGE

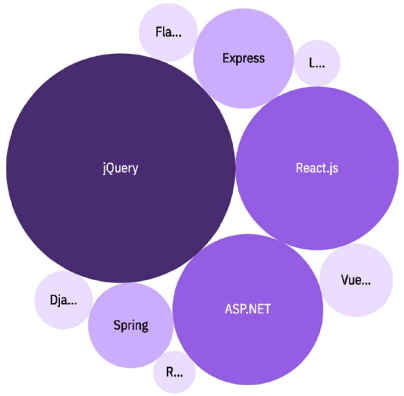
Top 10 LanguageWorkedWith 1



Top 10 DatabaseWorkedWith 2



Top 10 WebFramesWorkedWith 3



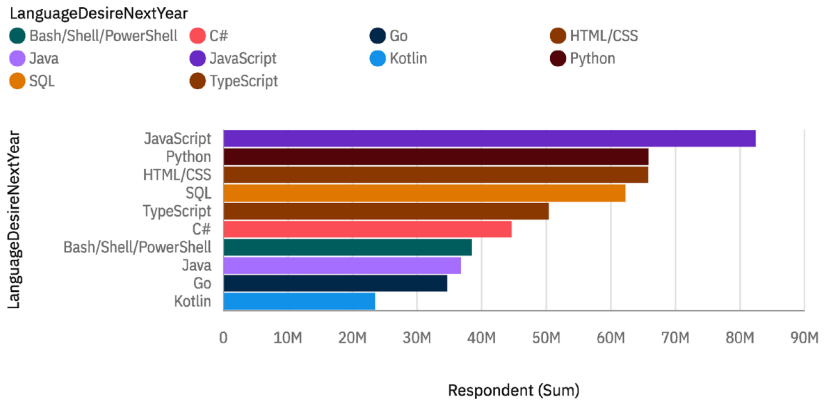
PlatformWorkedWith

DASHBOARD TAB 2

FUTURE TECHNOLOGY TRENDS

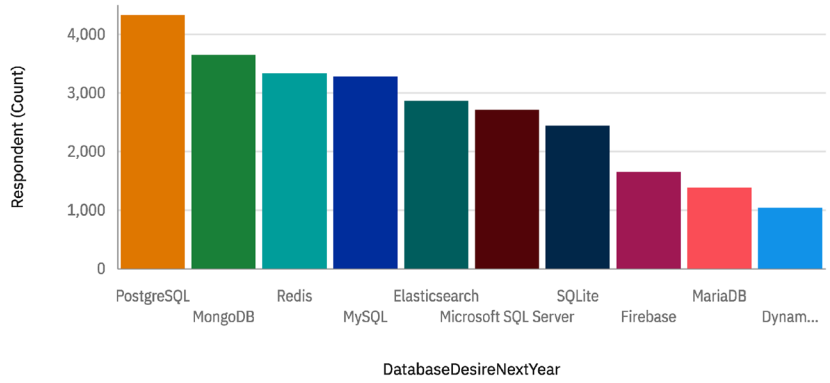
Top 10 LanguageDesireNextYear

4



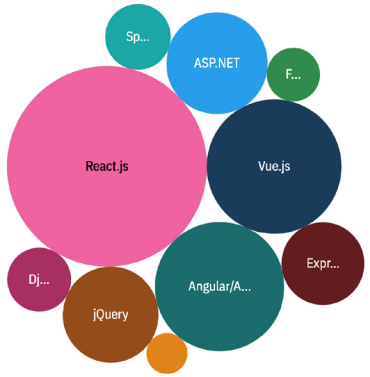
Top 10 DatabaseDesireNextYear

5



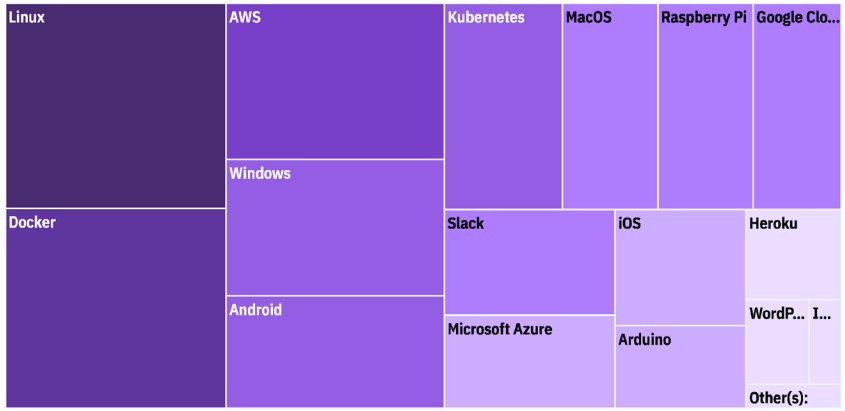
Top 10 WebFrameDesiredNextYear

6



PlatformDesireNextYear

7



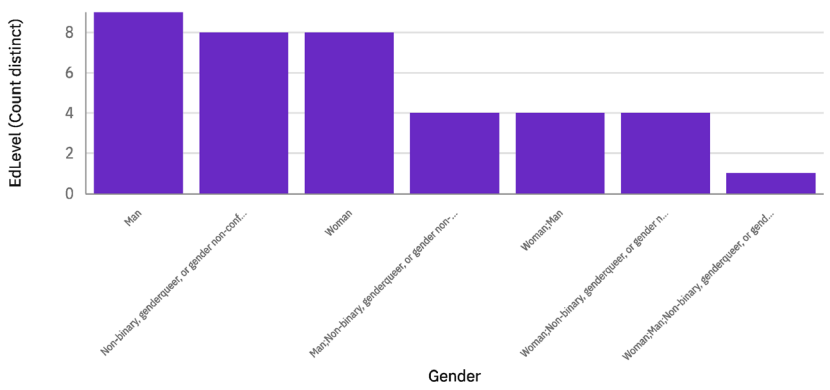
IBM Developer

SKILLS NETWORK

DASHBOARD TAB 3

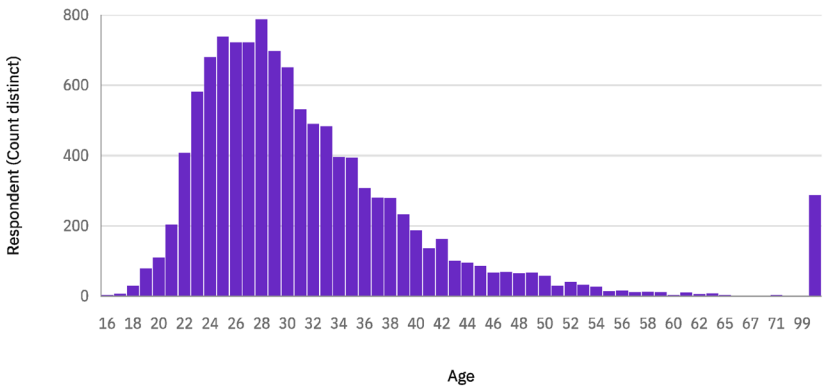
DEMOGRAPHICS

EdLevel by Gender

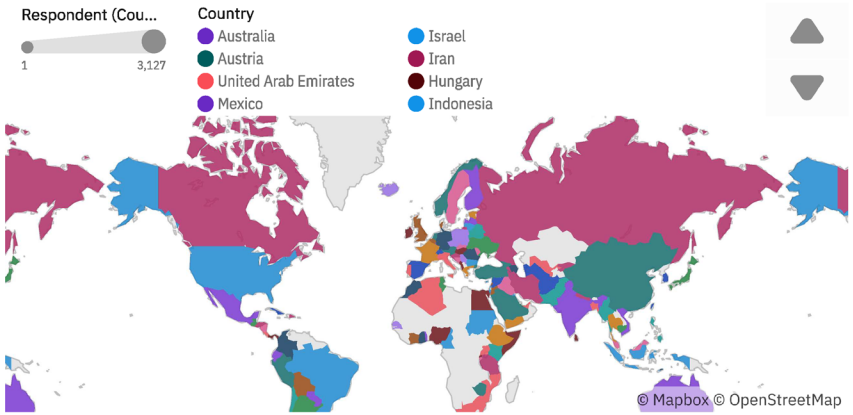


8

Respondents by Age

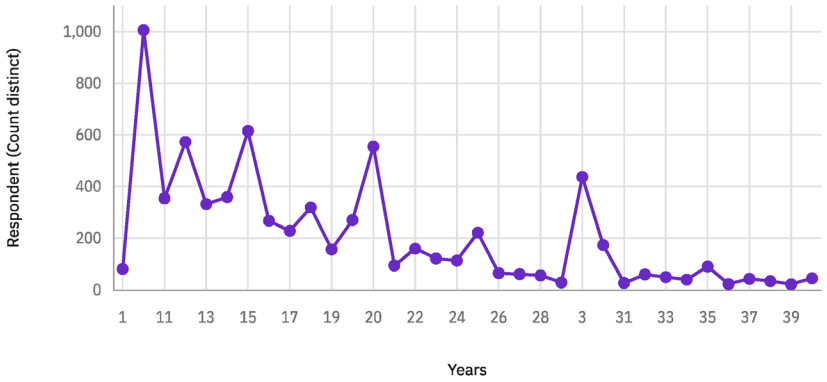


Respondents by Country



Respondent by YearsCode

9



DISCUSSION

TBD

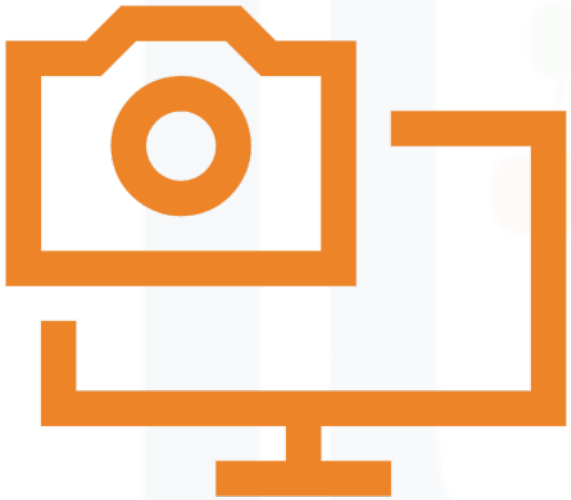
OVERALL FINDINGS & IMPLICATIONS

- **Programming Language Trends:**
  - **Findings:** Shifts in desired languages, with C#, C, and C++ rising in rank; new entrants like Go, Dart, and Assembly.
  - **Implications:** The evolving needs suggest a dynamic shift in industry requirements, indicating the need to stay updated with newer languages.
- **Database Trends:**
  - **Findings:** Notable changes include MySQL dropping in popularity and ElasticSearch rising; new entries like Cassandra and Couchbase.
  - **Implications:** As database preferences shift, companies need to reassess their database strategies to align with current trends.
- **Job Posting Analysis:**
  - **Findings:** Varying demand across regions and industries for specific languages and databases.
  - **Implications:** Understanding these trends can guide job seekers and companies in focusing on in-demand skills and technologies.

CONCLUSION

- **Methodology Strength:**
  - High participation and comprehensive data cleaning ensure reliable findings.
- **Survey Evolution:**
  - Shift from language/database focus to feature-based insights reflects changing industry needs.
- **Ranked Voting Impact:**
  - Enhanced feature prioritization through flexible voting system captures user preferences more accurately.
- **Industry Trends:**
  - Analysis of job postings and database trends highlights dynamic shifts in technology demand.

APPENDIX



- Include any relevant additional charts, or tables that you may have created during the analysis phase.

JOB POSTINGS TRACK OVER TIME

- **Trend Analysis:** Track the frequency of mentions over time to spot rising or declining demand.
- **Regional Insights:** Identify geographic hotspots for specific skills.
- **Industry Focus:** Determine which industries prefer certain languages and databases.
- **Data Collection:** Aggregate job postings mentioning specific languages and databases.
- **Skill Combinations:** Spot common pairings of languages and databases in job postings.
- **Visualization:** Use charts to show changes in demand and technology adoption.
- **Forecasting:** Predict future demand based on historical trends.

POPULAR LANGUAGES

In Module 1 you have collected the job postings data using web scraping in a file named “popular-languages.csv”. Present that data using a bar chart here. Order the bar chart in the descending order of salary.