Training for the Future

Tuan Duong

February 21, 2022

OUTLINE

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

EXECUTIVE SUMMARY

- IBM is looking to invest money to their employees for future technology training.
- This analysis will look at the present and future needs based on a survey.
  - Languages
  - Databases
  - Platforms
  - Web Frame
- Then it will look at the demographics for diversity.
- Lastly, a dashboard will be made for some EDA
- To conclude, we will talk about the insight.

INTRODUCTION

- This project is being presented in a role of a data analyst working for fictitious technology company, IBN.
- It was prepared for the IBM Data Analyst Capstone.
- This report includes the data analyst methodology, results, and conclusions on their findings.
- IBN is looking invest in their employees to train them on future technology.
  - This presentation will look at the current technology.
  - Then look at what the future needs are.

METHODOLOGY

- Data Collection
- Data Wrangling
- EDA: Data Visualization
- Dashboard
  - Current Technology
  - Future Needs
- Results

Data Collection

You can also view the json file contents from the following json URL.

```
In [92]: api_url="http://127.0.0.1:5000/data"
         r=requests.get(api_url)
```

```
In [112]: def get_number_of_jobs_T(technology):
              reference={"Key Skills":technology}
              r=requests.get(api_url,params=reference)
              df = pd.read_json(r.text)
              number_of_jobs = df.shape[0]
              return technology,number_of_jobs
```

Calling the function for Python and checking if it works.

```
In [113]: get_number_of_jobs_T("Python")
Out[113]: ('Python', 1173)
```

Write a function to find number of jobs in US for a location of your choice

```
In [114]: def get_number_of_jobs_L(location):
              reference={"Location":location}
              r=requests.get(api_url,params=reference)
              df = pd.read_json(r.text)
              number_of_jobs = df.shape[0]
              return location,number_of_jobs
```

- Data was collecting through an API
- Data was collected through web scraping
- Data was collected through a database

Data Wrangling

Finding Missing values

Find the missing values for all columns.

```
In [77]: # your code goes here
         #check for null
         df.isnull().sum()
```

```
Out[77]: Respondent        0
         MainBranch        0
         Hobbyist          0
         OpenSourcer       0
         OpenSource       81
                          ...
         Sexuality       542
         Ethnicity       675
         Dependents      140
         SurveyLength     19
         SurveyEase       14
         Length: 85, dtype: int64
```

Find out how many rows are missing in the column 'WorkLoc'

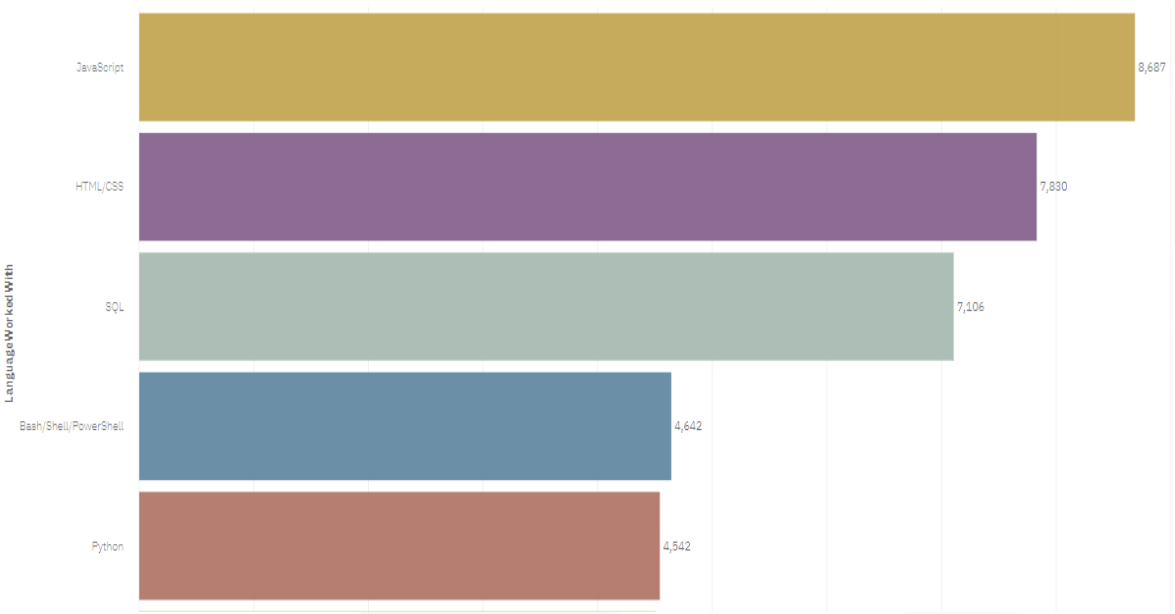
```
In [78]: df['WorkLoc'].isna().sum()
```

```
Out[78]: 32
```

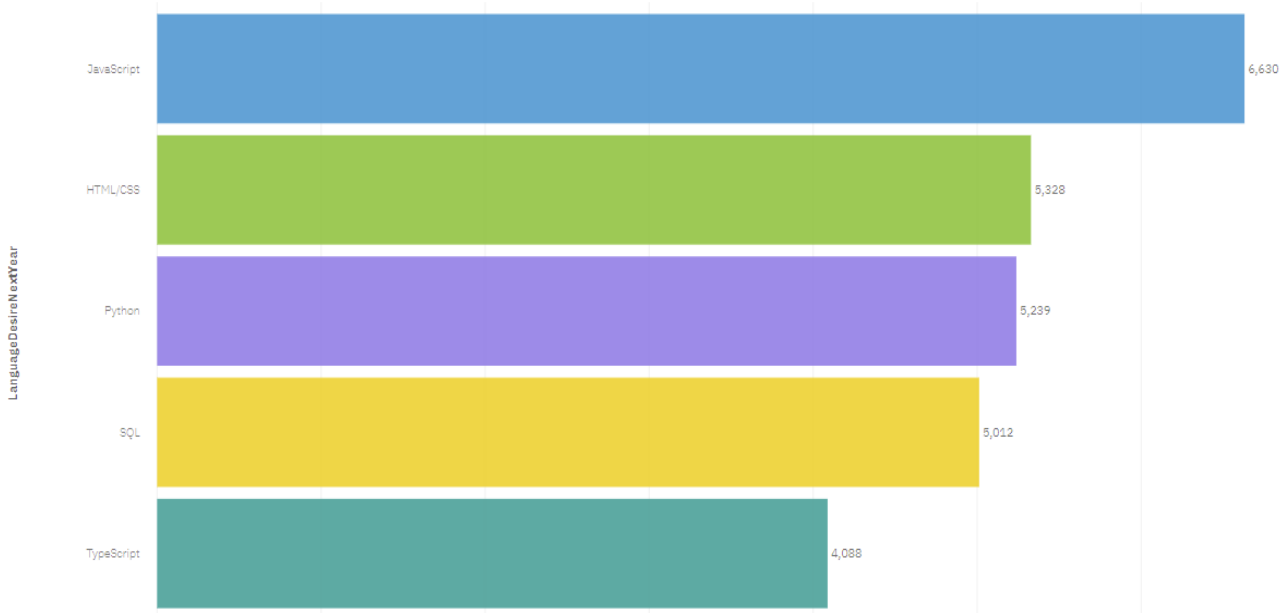
- Data was cleaned with the following techniques
  - Look for duplicates
  - Drop duplicates
  - Look for Null values
  - Impute Null values
  - Look for outliers
  - Drop outliers
  - Normalized Data

EDA: PROGRAMMING LANGUAGE TRENDS

Current Year



Next Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

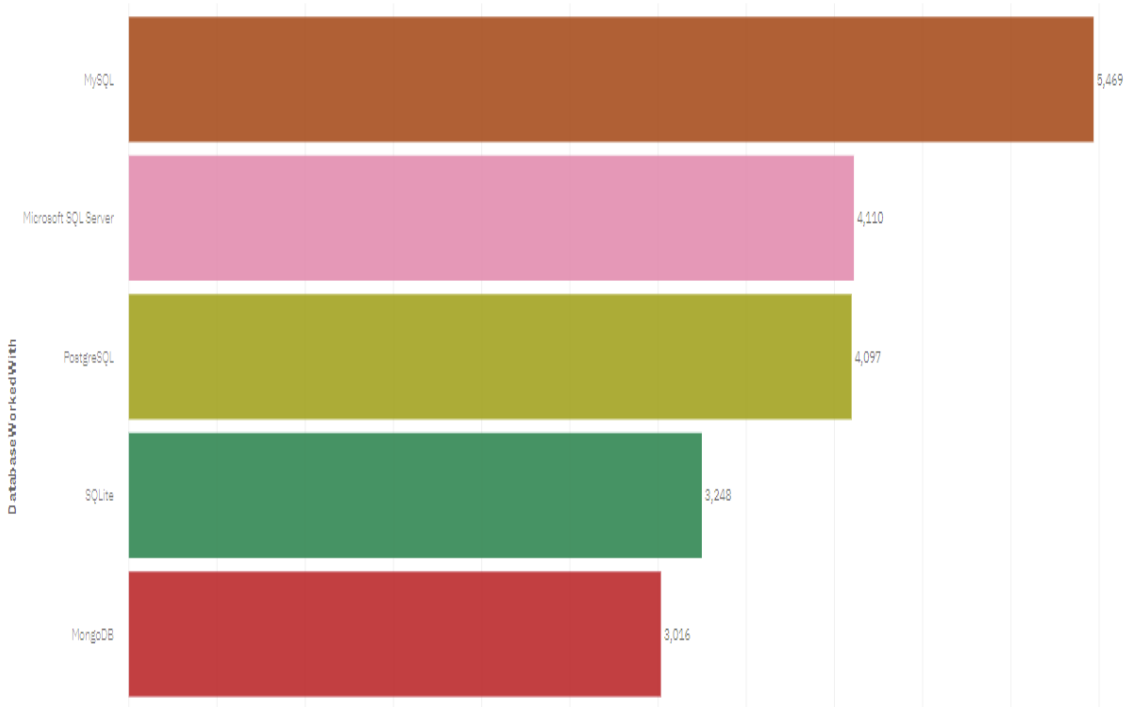
- Java Script is still most desired.
- HTML is second.
- Python is upcoming.

Implications

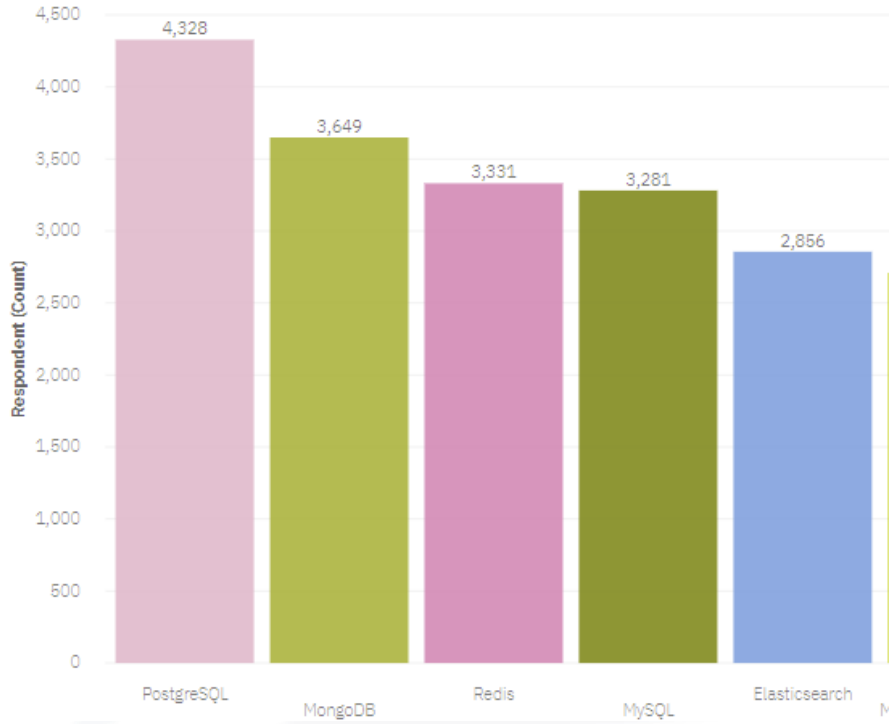
- Should still invest in JavaScript
- Should still invest in HTML
- Look into Python/SQL next year.

EDA: DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- PostgreSQL is most desired
- Followed by MongoDB
- Followed by Redis

Implications

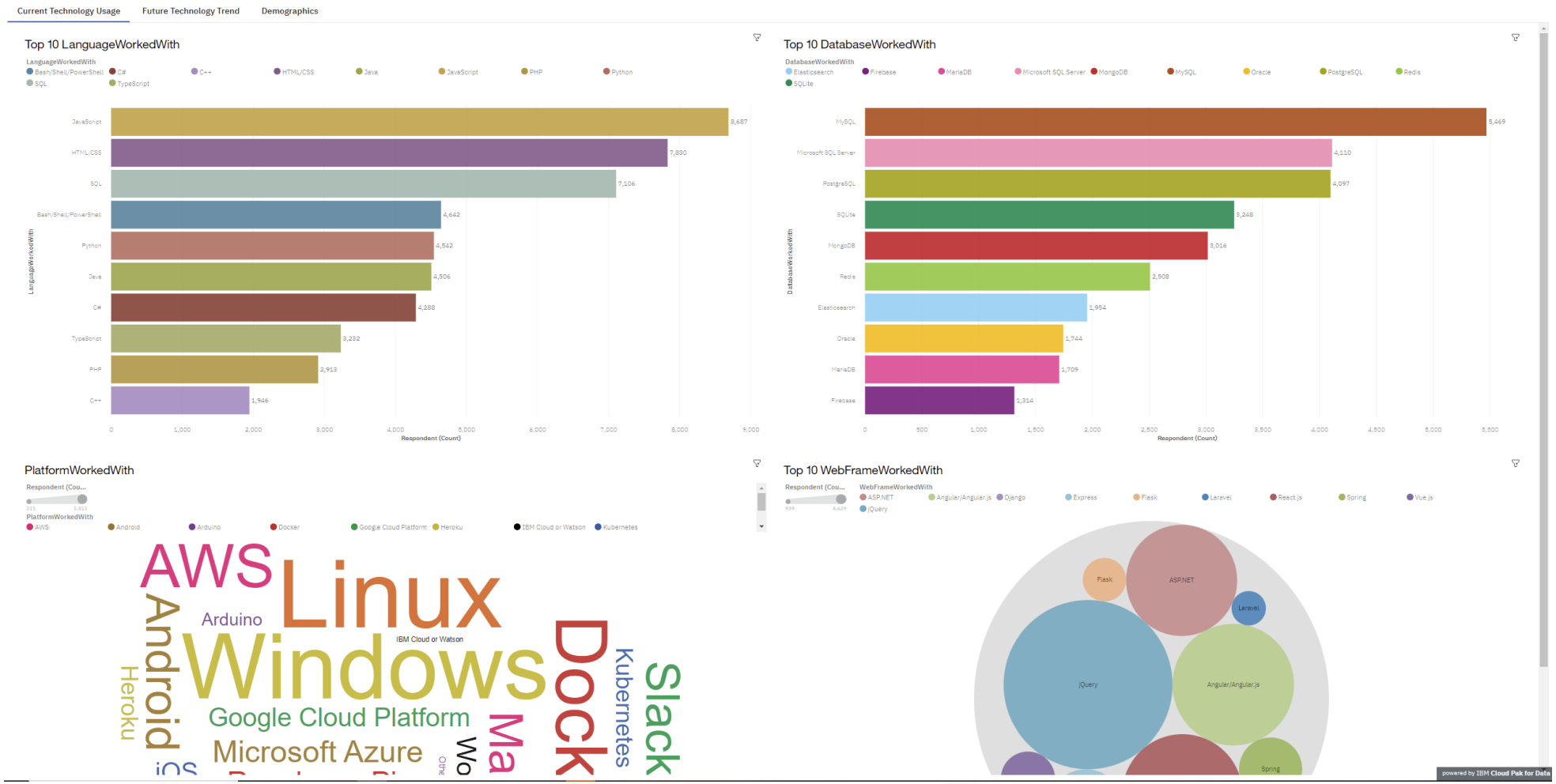
- Should invest in PostgreSQL
- Maybe invest in MongoDB
- Look for other upcoming databases next year.

DASHBOARD

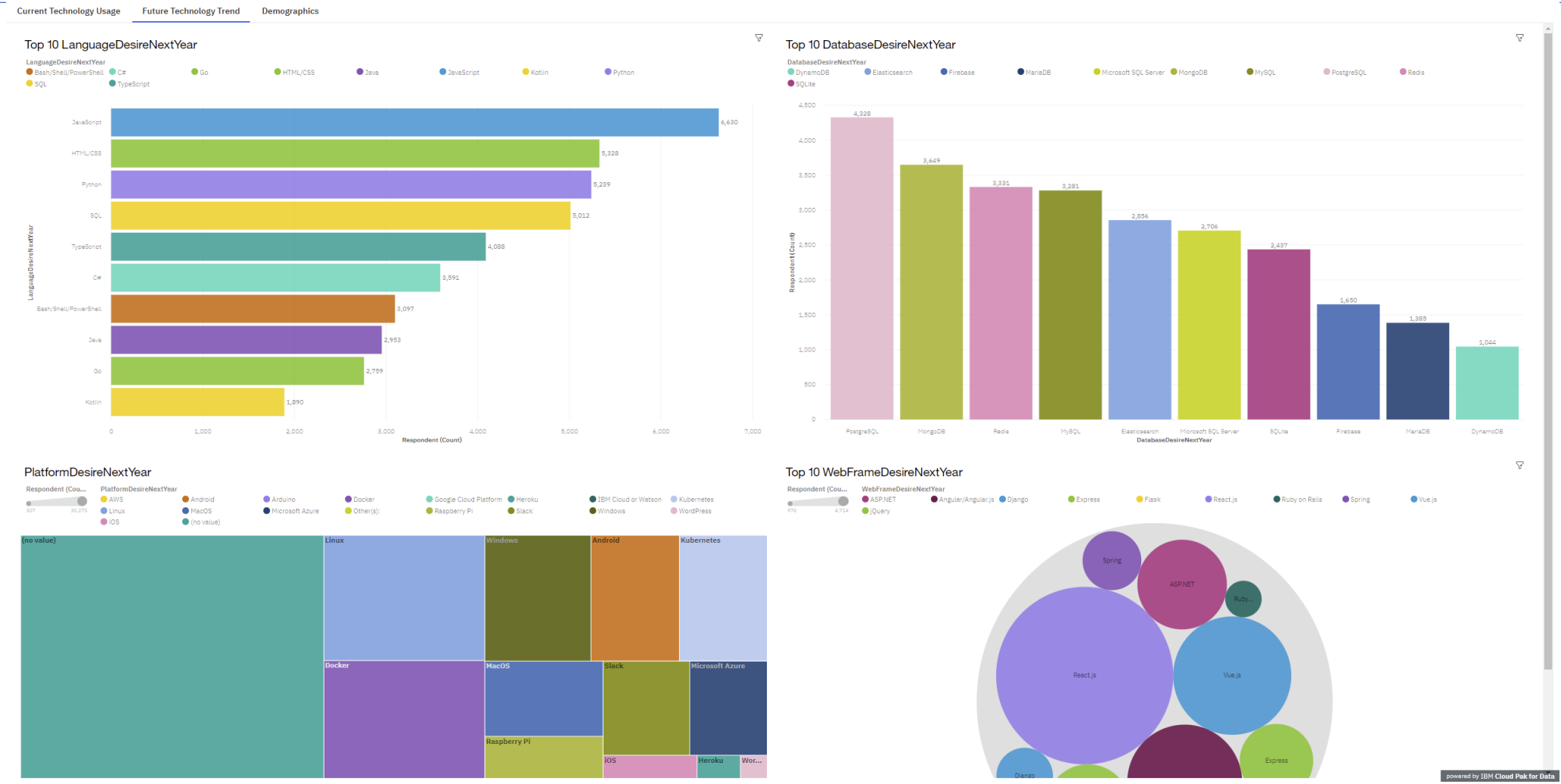
Link to Dashboard

https://dataplatform.cloud.ibm.com/dashboards/6660ed24-2bfd-4555-b641-9cb236d2730c/view/6333c301339903ff6be9f2e4079078537e662c08b5bb810ad68d7b4907317997a93a1294c82e1908d8400667fabd155a98

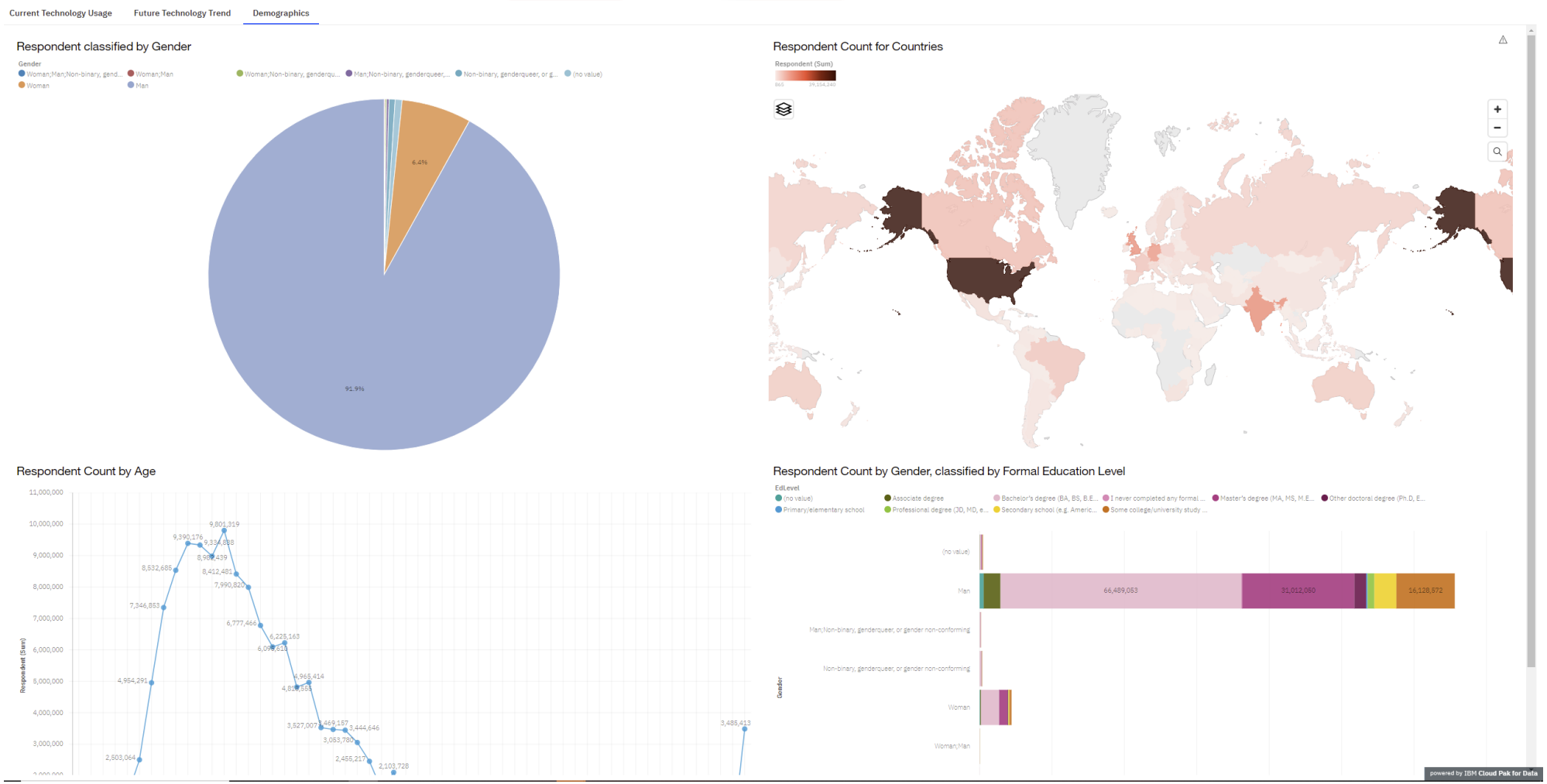
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



IBM Developer

SKILLS NETWORK

OVERALL FINDINGS & IMPLICATIONS

Findings

- Language developing is at a good spot
- Database developing needs to be updated
- Diversity may need to be looked at

Implications

- No Change in Language
- Should invest in new database training with PostgreSQL
- Diversifying might help bring in different views

RESULTS

- Results show that a shift technology may be trending
- Research committee should be formed for technology that should be updated

DISCUSSION

- Further research needs to be done for platform and web frame future technologies
- One indication is database, web frame network, and platform technology may be changing
- Board should discuss where is the best area to update

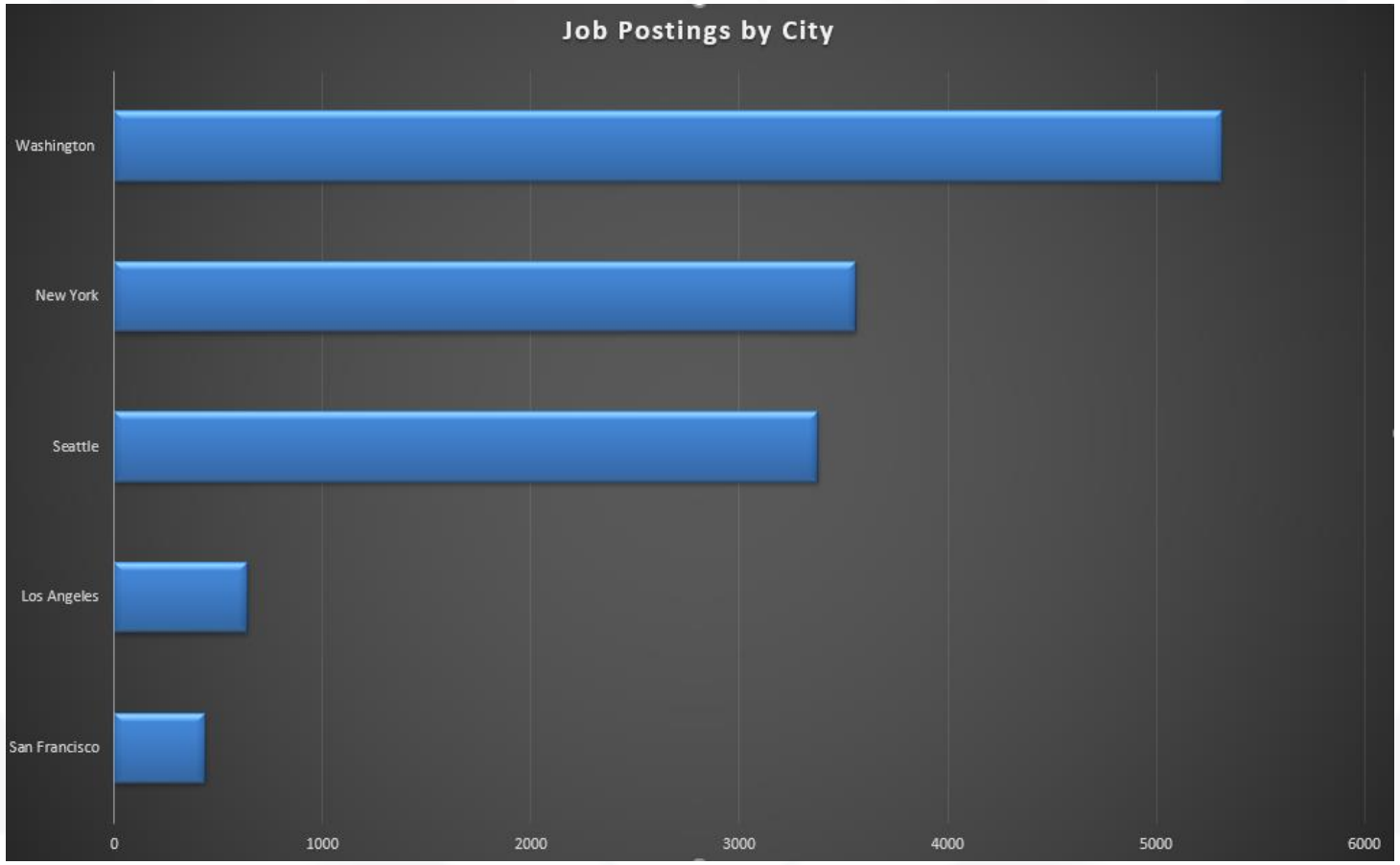
CONCLUSION

- Data Analyst performed data acquisition, wrangling, and validation
- Data Analyst went through an EDA through SQL, Data Visualization, and Dashboard making
- Data Analyst looked through analysis for insight
- Insight was found through a shift of technology needs.

APPENDIX

- The project was created by the following applications: Python, Jupyter Notebook, SQLLite, IBM Cognos Microsoft Excel, Microsoft Power Point. These libraries were used in Python: Pandas, Numpy, Matplotlib, Seaborn, Requests, Beautiful Soup, API, and Wget.

GITHUB JOB POSTINGS



POPULAR LANGUAGES