FINAL PRESENTATION IBM DATA ANALYST CAPSTONE

MANOJ V

15/01/2023

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

EXECUTIVE SUMMARY

- The Technology in modern world is constantly evolving and it is a necessity to upskill regularly
- To keep up with evolving technology and to stay competitive, it's critical to understand future skill requirements and trends.
- This presentation will show current and future trends in Programming Languages, Databases, Platforms and Web frames.
- The presentation will also aid in recognizing key skill shortages and the analysis will present data that can be used to make more informed data-driven budgetary decisions.

Introduction

- Objective: To Intepret the survey data from stackoverflow and use it for analysis of current and future trends
- Desirable Outcomes:
- Show strong correlation in key skill requirements data using SQL and Python.
- Dashboard analytics to show key charts and maps.

METHODOLOGY

- Extracting Data from API from python
- Web scrapping using Beautiful Soup Package
- Data Wrangling using Pandas
- Exploratory Data Analysis using Seaborn and Matplotlib
- Data Visualization for the findings
- Dashboard in IBM Cognos

Methodology - Explained

- Extracting Data from API from python - The data collection is carried through an API that connects to the survey data. URL: http://127.0.0.1:5000/data
- Web scrapping using Beautiful Soup Package - Beautiful Soup object is created, and data is extracted. Data frame is used to create columns Language and Avg Annual Salary
- Data Wrangling using Pandas – Finding Duplicate rows and removal, finding and filling missing values and normalization using pandas functions
- Exploratory Data Analysis using Seaborn and Matplotlib – Understanding the data using Python libraries
- Data Visualization for the findings – Visualizing the analysis to check for consistency
- Dashboard in IBM Cognos – create dashboards for publishing the results

RESULTS - EXTRACTING DATA FROM API FROM PYTHON

```
Locations = ["Los Angeles", "New York", "Washington DC", "San Francisco", "Detroit", "Seattle", "Austin"]

Result = []

def get_number_of_jobs_Locs(Locations):
    for i in Locations:
        payload = {"Location": i}
        response = requests.get(api_url, params=payload)
        if response.ok:
            data = response.json()
            number_of_jobs = len(data)
            Result.append([i,number_of_jobs])
    return Result
```

```
get_number_of_jobs_Locs(Locations)
```

```
[['Los Angeles', 640],
 ['New York', 3226],
 ['Washington DC', 5316],
 ['San Francisco', 435],
 ['Detroit', 3945],
 ['Seattle', 3375],
 ['Austin', 434]]
```

RESULTS - WEB SCRAPPING USING BEAUTIFUL SOUP PACKAGE

```
Soup = BeautifulSoup(data,'html')
```

Scrape the `Language name` and `annual average salary`.

```
table = Soup.find("table")
language_name = []
Average_Salary = []
for row in table.find_all("tr"):
    language_name.append(row.find_all("td")[1].getText())
    Average_Salary.append(row.find_all("td")[3].getText())

df = pd.DataFrame(columns=[language_name,Average_Salary]).T.reset_index()
df = df.rename(columns={"level_0": "Language", "level_1": "Average Annual Salary"})[1:]
```

	Language	Average Annual Salary
1	Python	$114,383
2	Java	$101,013
3	R	$92,037
4	Javascript	$110,981
5	Swift	$130,801
6	C++	$113,865
7	C#	$88,726
8	PHP	$84,727
9	SQL	$84,793
10	Go	$94,082

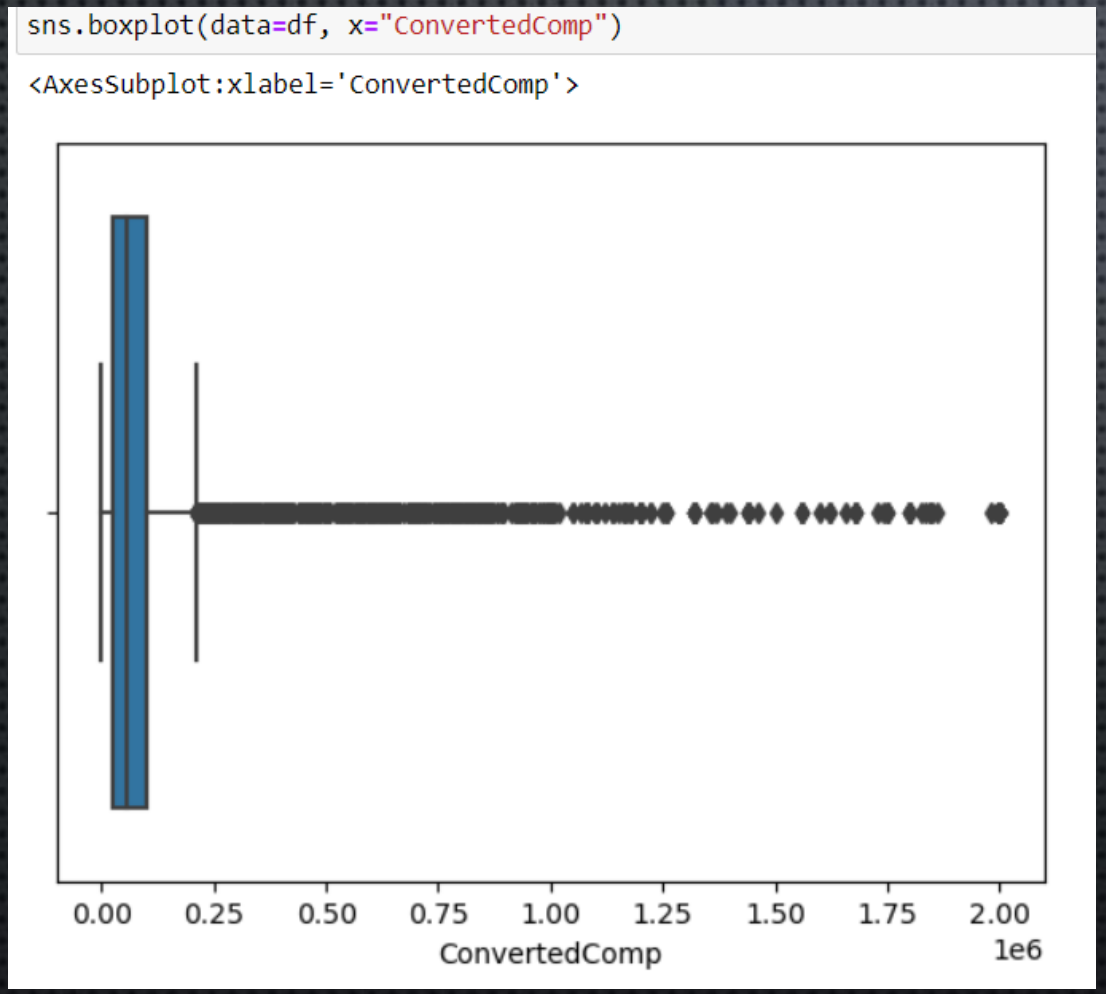
```
for x,y in zip(df['CompFreq'], df['CompTotal']):
    if x=='Monthly':
        comp.append(y*12)
    elif x=='Weekly':
        comp.append(y*52)
    else:
        comp.append(y)

df['NormalizedAnnualCompensation']= comp
df[['NormalizedAnnualCompensation']]
```

	NormalizedAnnualCompensation
0	61000.0
1	138000.0
2	90000.0
3	348000.0
4	90000.0
...	...
11547	130000.0
11548	74400.0
11549	105000.0
11550	80000.0
11551	NaN

RESULTS - DATA WRANGLING USING PANDAS

```
sns.boxplot(data=df, x="ConvertedComp")
```

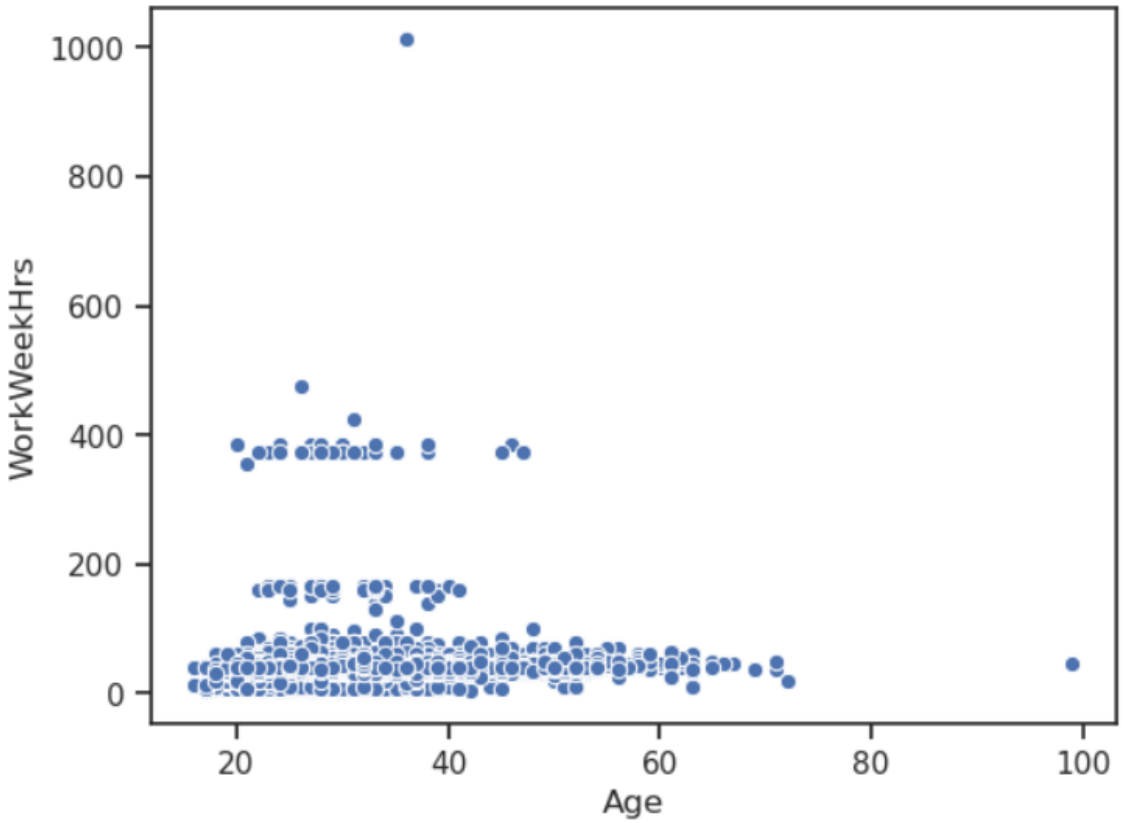
```
<AxesSubplot:xlabel='ConvertedComp'>
```



RESULTS - EXPLORATORY DATA ANALYSIS USING SEABORN AND MATPLOTLIB

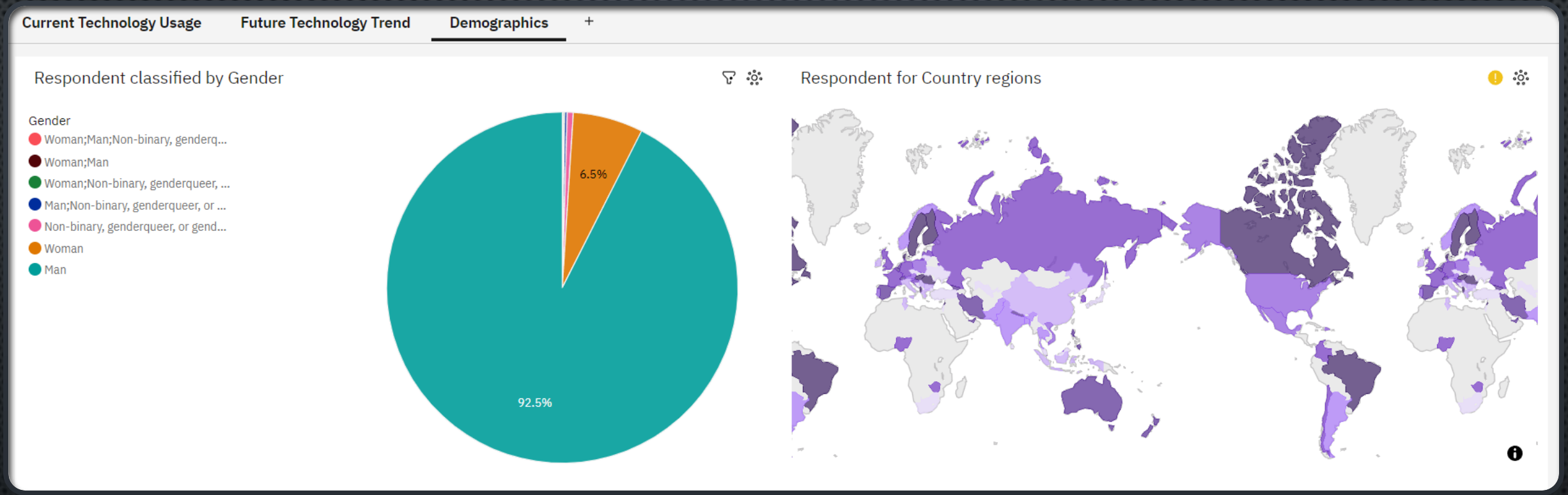
```
Query = """
Select WorkWeekHrs, Age from master"""

WorkWeekHrs = pd.read_sql_query(Query, conn)

sns.scatterplot(data = WorkWeekHrs, x = "Age", y = "WorkWeekHrs")
```

```
<AxesSubplot:xlabel='Age', ylabel='WorkWeekHrs'>
```



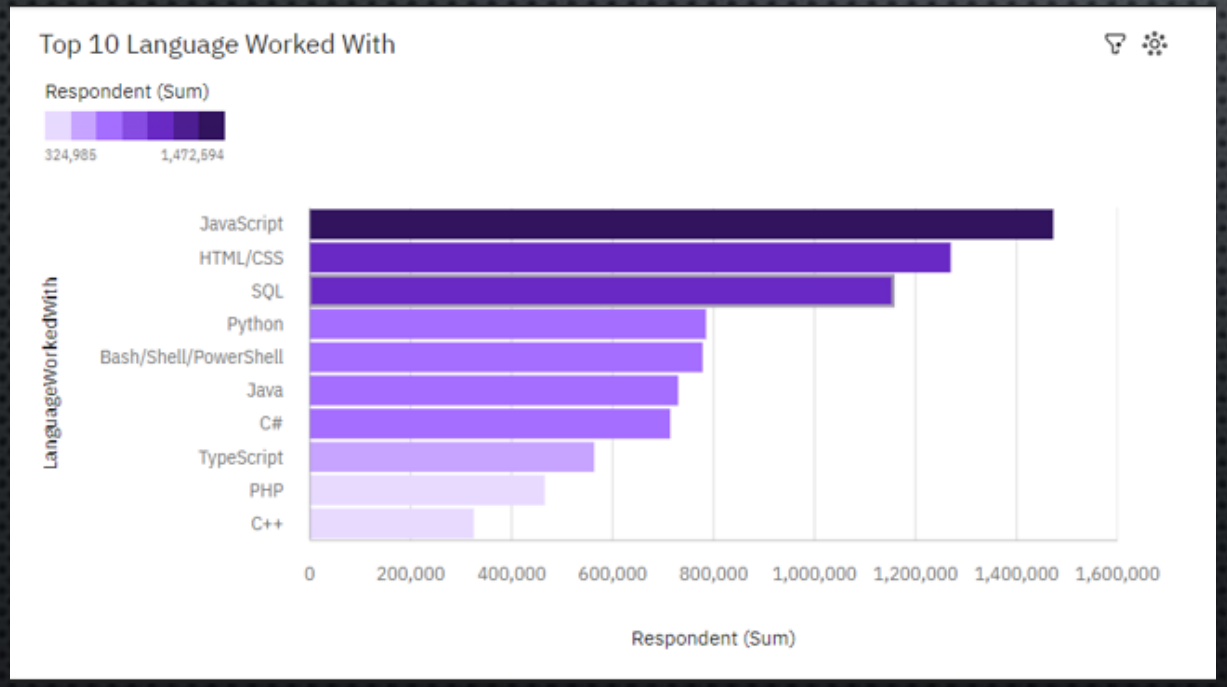
RESULTS - DATA VISUALIZATION FOR THE FINDINGS

RESULTS - DASHBOARD IN IBM COGNOS

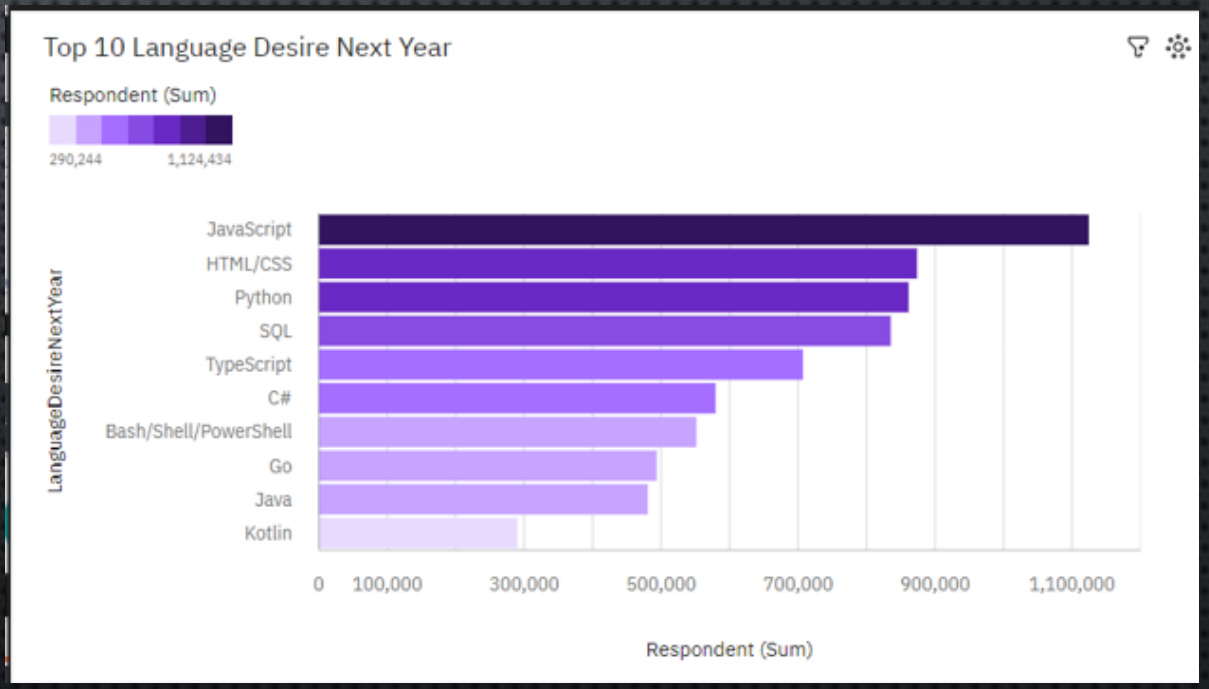


PROGRAMMING LANGUAGE TRENDS

CURRENT YEAR



NEXT YEAR



Programming Language Trends - Findings & Implications

Findings

JavaScript, HTML/CSS continue to be the most sought after languages in the future and most desired

Python is 3rd most desired language in the next year

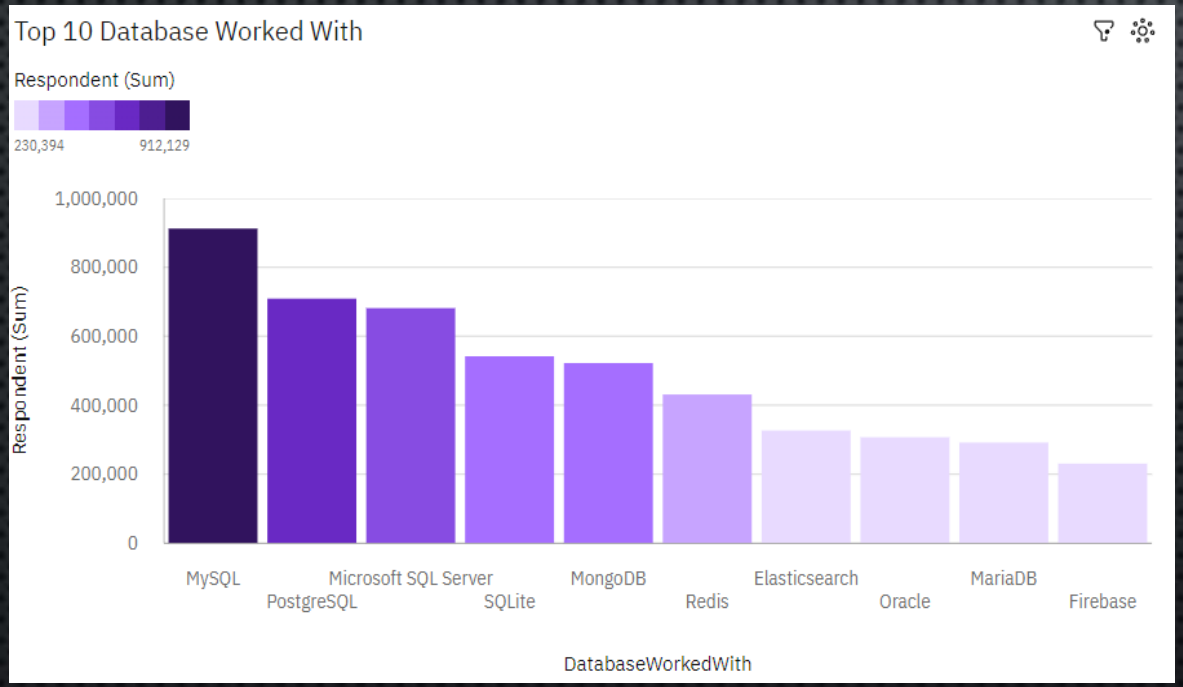
Implications

Employment with JavaScript and HTML/CSS will remain in demand next year.

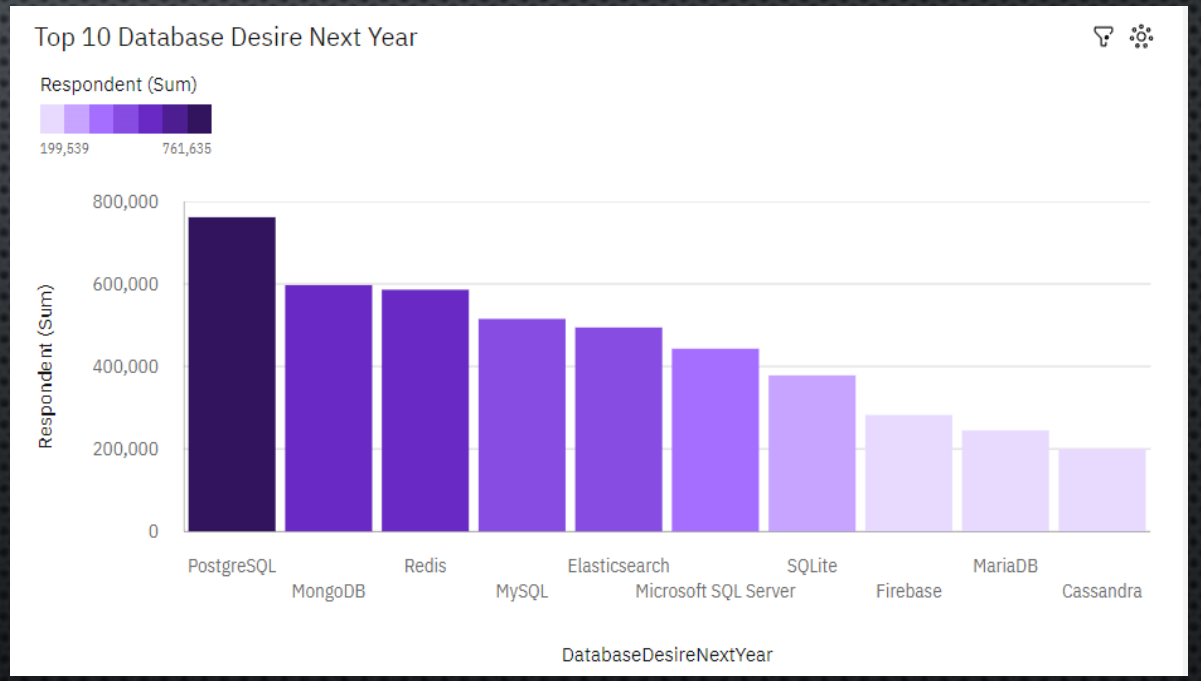
Python developers employment demand will surge next year.

DATABASE TRENDS

CURRENT YEAR



NEXT YEAR



Programming language trends - findings & implications

Findings

PostgreSQL has risen in demand next year

MySQL demand dips in the following year

Implications

Non relationship database languages have seen an increase and PostgreSQL improved the most

MySQL openings might dip in the future

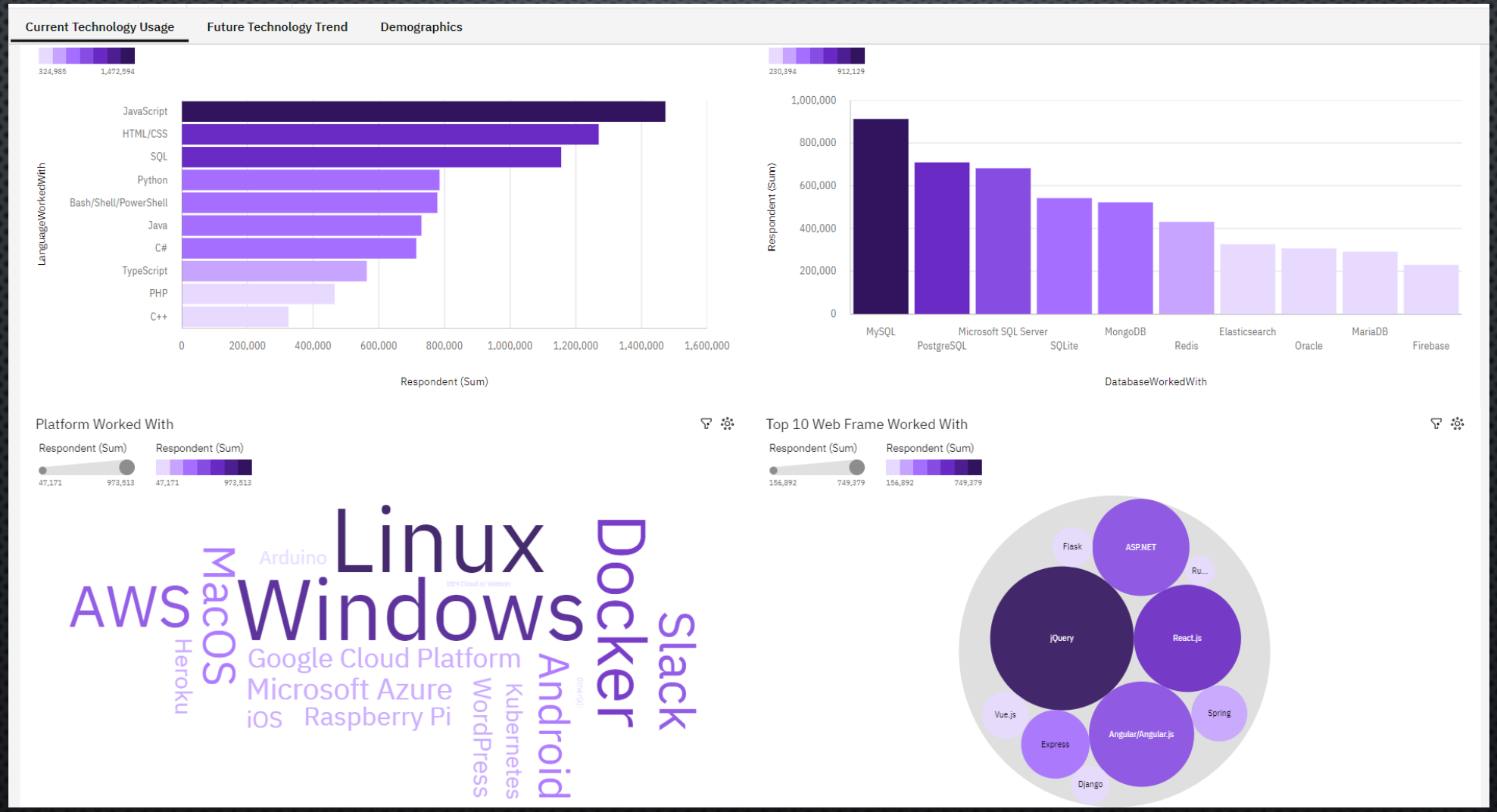
DASHBOARD

Link for Cognos Dashboard:

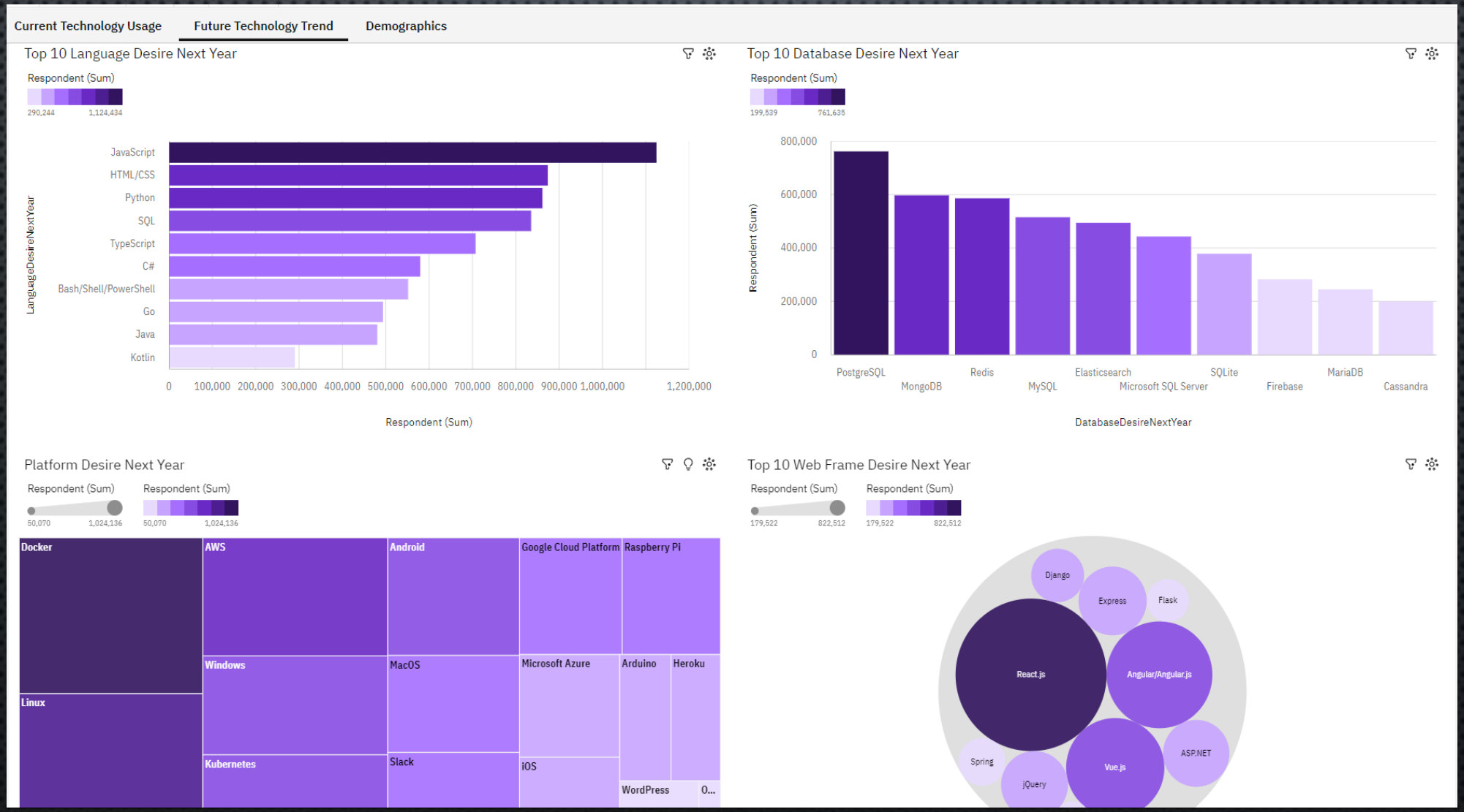
https://eu2.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FM5&action=view&mode=dashboard&subView=model00000185b0d30d17_00000000

(Login with IBM Account before opening the link)

DASHBOARD TAB 1



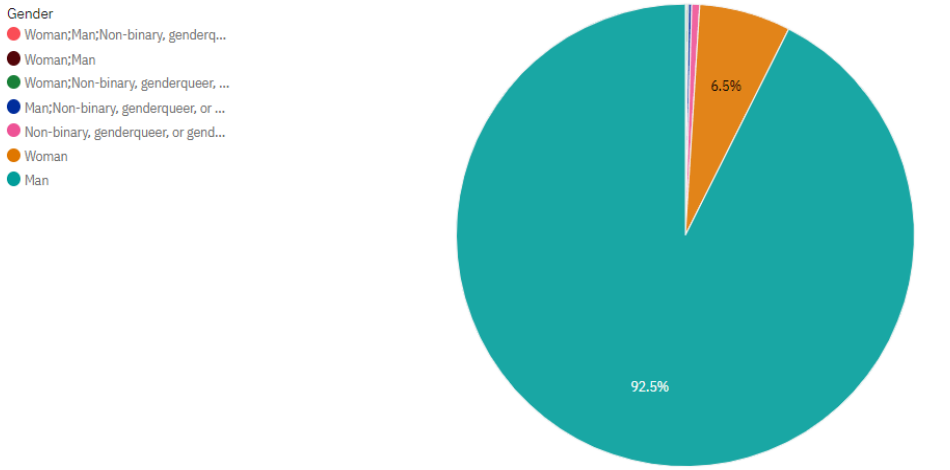
DASHBOARD TAB 2



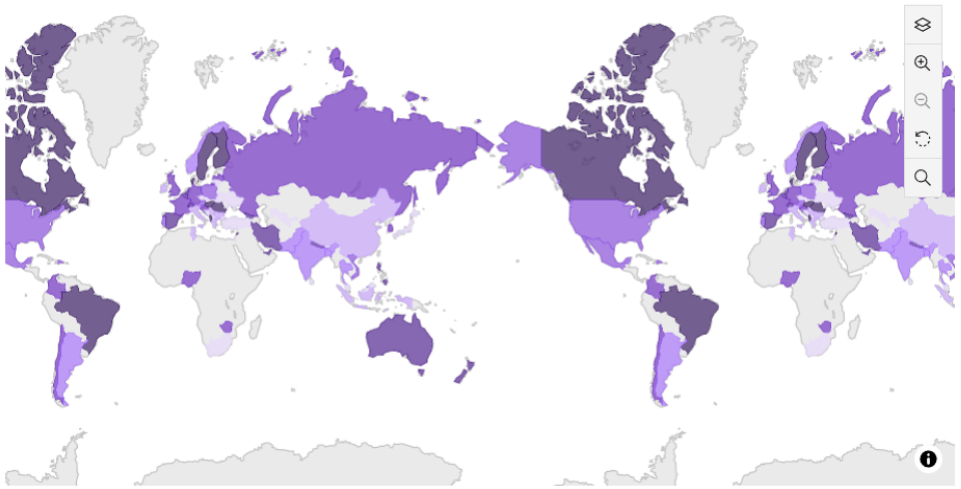
DASHBOARD TAB 3

Current Technology Usage | Future Technology Trend | Demographics

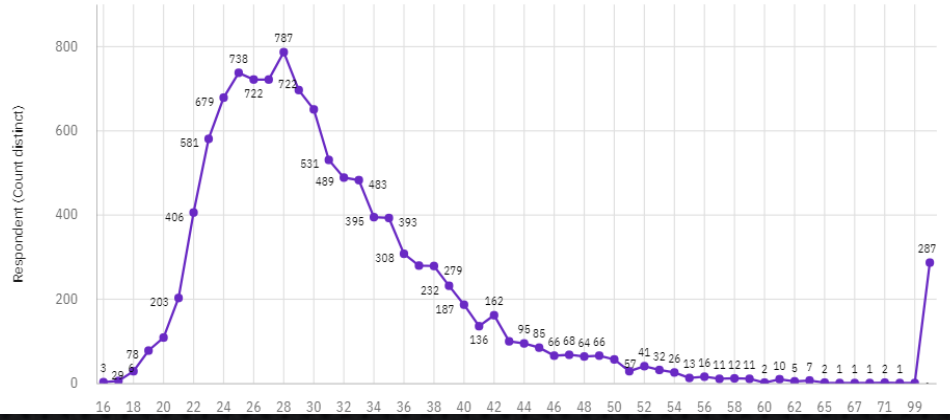
Respondent classified by Gender



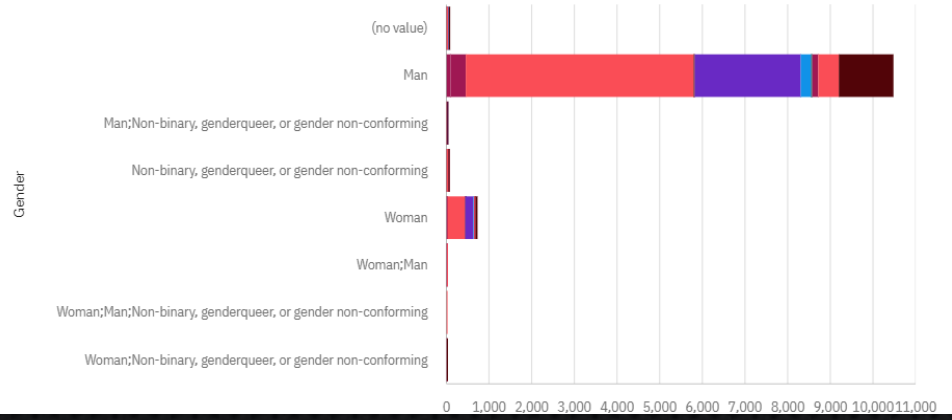
Respondent for Country regions



Respondent Count by Age



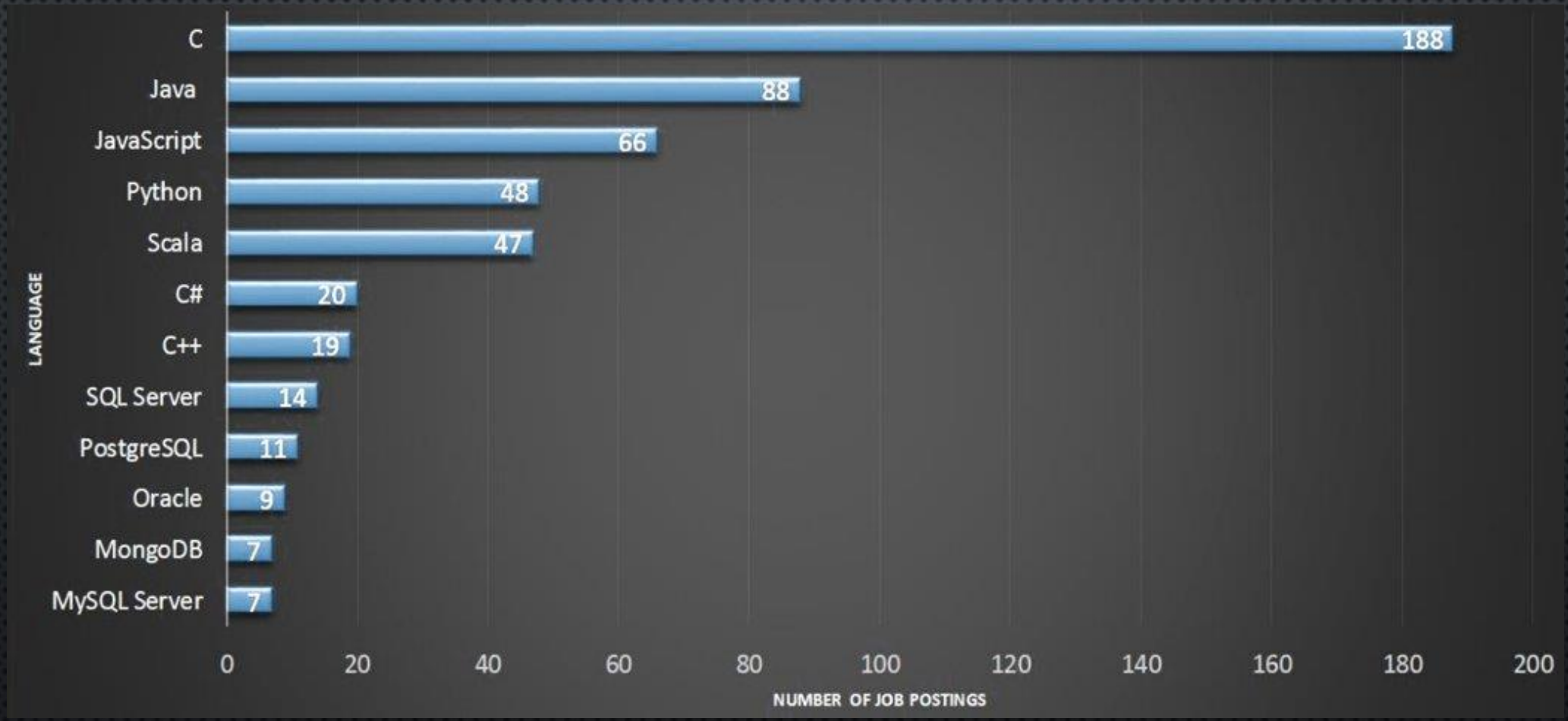
Respondent Count by Gender, classified by Formal Education Level



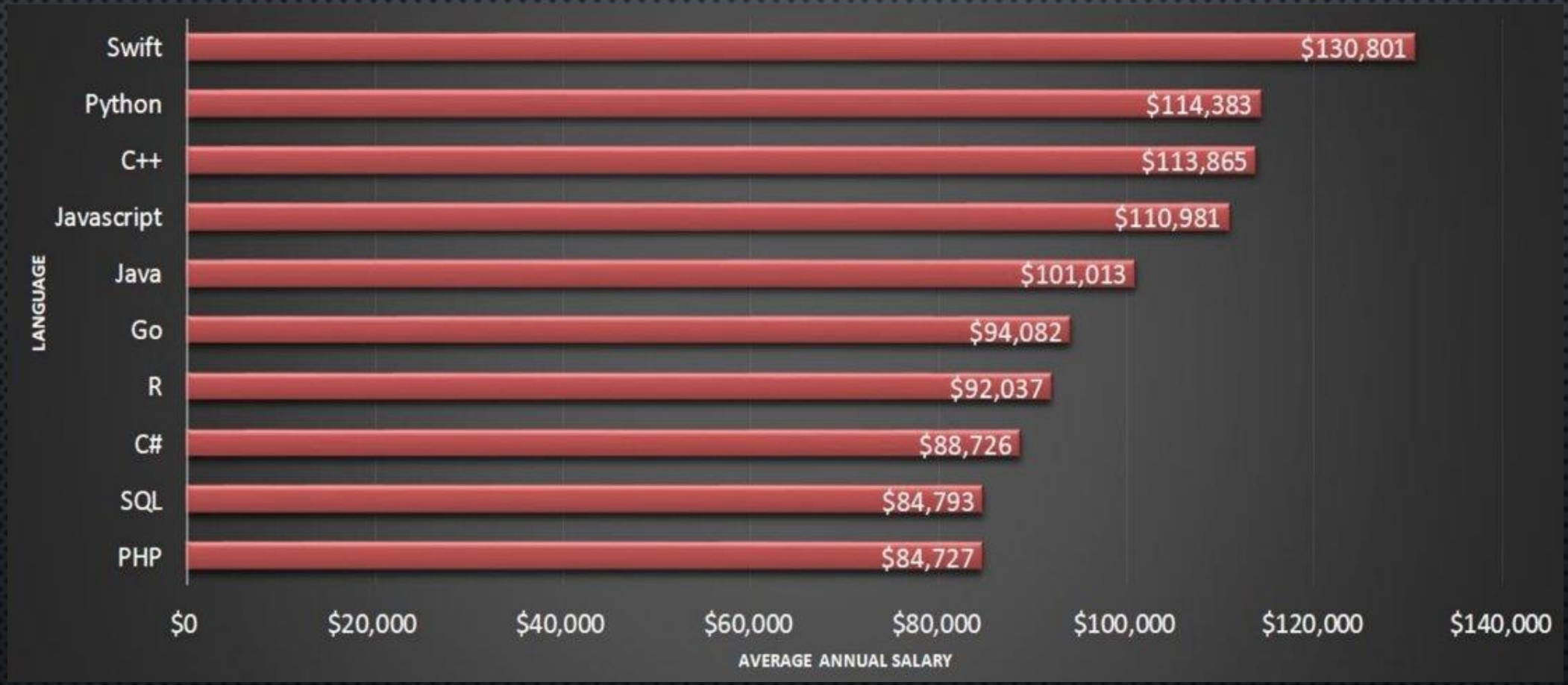
DISCUSSION

1. The representation of women is very less comparatively which should be addressed in the future
2. Certain Programming Languages like Python and Databases like MongoDB have a rise in popularity amongst developers due to flexibility and ease of use
3. Upskilling is an mandatory task done to keep employees up to date with the latest trend.
4. Companies should assign few hours for employees to work on upskilling their skillset regularly

JOB POSTINGS



POPULAR LANGUAGES



THANK YOU FOR YOUR TIME