Project Report

**IT Salary Survey of Europe 2020**

IS 680

Data Science I

Presented by:

Dhananjay Panchal

Dhaval Vibhakar

Mohammed Ismael

Satyajit Kadam

Mohamed Zaki

Under the Guidance Of

Prof. Yegin Genc

Index

|  |  |  |
| --- | --- | --- |
| Chapter No | Topic | Page No |
| 1 | INTRODUCTION | 3 |
| 1.1 Description |
| 1.2 Problem Formulation |
| 1.3 Proposed Solution |
|  |
|  |
| 2 | Data   * 1. Source   2. Data preprocessing   3. Descriptive Analysis | 4 |
| 3 | Exploratory Analysis | 5 |
|  |
|  |
| 4 | Conclusion | 13 |
|  |

**Introduction**

* 1. Description

The year 2020 is one that not even one of us will presumably at any point neglect, and the effect that the pandemic had on our positions and vocations truly fluctuated relying upon industry and occupation work. Certain individuals were more occupied than at any other time, and others saw their work hours cut back or their positions wiped out. IT was vital to aiding their associations turn to advanced business and exchanges and aiding office laborers set ready for telecommuting.

* 1. Problem Formulation

Nowdays finding a job with the average pay is too difficult and many fresher/employees are still unclear with their positions. In addition, Information Technology is a vast field and has no stop to it as this field also has it importance and many other field rely on it too. We also need to find the which gender-wise distribution of the jobs in a particular city. As there are many new companies uprising in the market with the older companies ,we need to analyze the salary of position in both these types.

* 1. Proposed Solution

This survey will guide numerous freshers and graduates to get an overview regarding which position ,role and software gets paid more and less and accordingly they can pave their path. It can also help many companies to work on their organisation with the reference to salary.

**Data**

* 1. Source

The dataset was found on Kaggle and the survey involves data from European country like Germany.

* 1. Variables

There are total 1253 rows and 23 columns. The variables used in exploratory analysis are Position, Gender , Seniority level, Programming Language ,Company type.

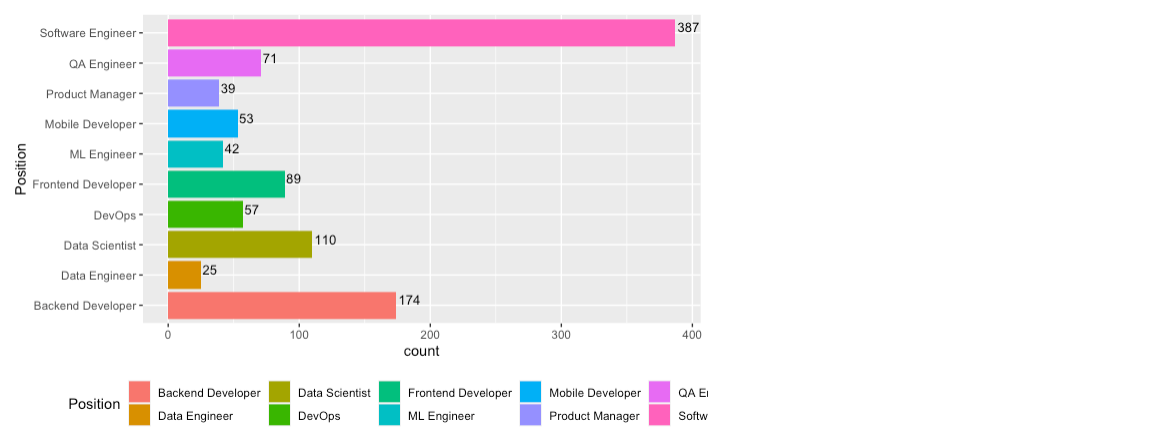
Data Cleaning

* Data Cleaning involves check for missing values as the dataset had many null values in numerical as well character based columns .
* We also had renamed columns inorder to run the columns accordance with the R guidelines
* The Column names where lowercase so that the libraries can read the table.
* Few columns had different data type so according to their values we had rearranged the whole datatype of the dataset.
  1. Descriptive Analysis

There are 19 characters and 4 numeric values. There are various variable types like Age has a mean of 32.5 and standard deviation 5.66.Age has a standard deviation of 5.66. The yearly salary in Eu has a mean of 80279043.The annual salary has a mean of 1.01e and other numeric value lie in the range of 1 to 6 as there are few missing values and data is cleaned. We observe that median has specific values.

**Exploratory Analysis**

1. Which Engineer is the most Dominating?



The Job position analysis help us to understand that Software Engineer is the most Dominant Position throughout Germany. We analysed that Position is an important value in our table and many relations can be analysed further in the project using this value.

1. Who has the highest Salary?

Chart, timeline, bar chart

Description automatically generated

In accordance with the graph as we analyse the salary in respective with the Job Position we found that Manager position has the highest paid salary with the value of 250000.

1. Which programming language is preferred by most IT Engineers?

A picture containing chart

Description automatically generated

As we see in the graph most IT engineers prefer Java as their preferred programming language and then python and javascript. This helps us to understand programming language demand in Germany.

4)Which languages are most preferred by Software engineer, Backend Developer, Data Scientist?

1.Software Engineer

Table

Description automatically generated

2.Back End Developer

Table

Description automatically generated

3.Data Scientist

Table

Description automatically generated

As we analysis the above the top engineering position, we find that Java and Python is mostly preferred by Software Engineers, Backend Developers and Data Scientist

5)Which are the other programming languages most used ?

Chart, bar chart

Description automatically generated

As we seen in the above graph the most preferred programming language is Docker, Python and sql.

6)Which role has highest number of full time employees, Part time employees and Self Employeed ?

Chart, pie chart

Description automatically generated

As we see Software Engineer are the highest number of full time employees then other types of engineers in Germany.

* Part Time employees

Chart, pie chart

Description automatically generated

Backend Developers have the highest number of employees working as part time in Germany.

* Self-Employeed

Chart, pie chart

Description automatically generated

As we can see from the above three analysis we observe that Software engineers have the most highest role in Full time employeed and Self Employeed and Back end Developer as Part Time employee.

7)Which Full-Time employee role has highest salary?



Machine Learning Engineer are the highest number of full time employed salaries then Product manager and others.

8)Which city has the highest number of survey taken and which position is leading in the highest surveyed city?

Chart, bar chart

Description automatically generated

Berlin has the highest number of highest survey taken and as we see below Male have taken the survey more than the females.

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

As we see from the results,Berlin as the highest number of survey taked in which the number of survey taken by male is more than the female.In Berlin the Software engineers are the highest leading position found.

9) What age-group are the employees belong to Software Engineer Position to see if there is any diverse age groups and which type of campnay do they prefer to work in?

Chart

Description automatically generated

We found out that median age group for this Position seems to be in range to 30-35 Age group with their Average salary as 62K$

Chart, bar chart

Description automatically generated

With above analysis we found out that every employee in the range of 30-35 having an average salary of $62k prefers a Product based company more then the startup based.

10) Which company type has the highest number of employees and has the highest salary?

Chart, bar chart

Description automatically generated

According to analysis as we see the Product based type of company have the highest number of employees working in Germany .

Chart, bar chart

Description automatically generated

As we see from both product based type has the highest amount of salary being paid in Germany.

11) In which type of company does the data scientist has the highest income ?

Timeline

Description automatically generated

As we analysis Data Scientist have the highest income in Consumer goods and then Coporate incubrators and others.

**Conclusion**

* In conclusion, we analysed that Software engineers plays a dominant role and leading position in the perspective with salary and also the number of surveys taken is more by them.
* Backend Developers are more in part time employment.
* Median age of 30-35 have highest paid salary.
* Product based company have more dominance than other type of companies in the market.