**ANALYTIC TOOLS & DECISION MAKING**

**Project Submission 1**

**PROJECT PROPOSAL**

Logo

Description automatically generated

**KNOW YOUR WORTH!**

**The Invincibles team**

* Rithesh Reddy Alwala
* Hemant Chowdary Gorantla
* Mounika Loka
* Aniketh Reddy Nayini
* Sreenatha Reddy Ramireddy
* Srujana Reddy Thangadipally
* Kiran Kumar Thutukuru Babu

**Introduction**

                       The German economy has always been a dynamic one. Germany has been recognized as a leader in digital and technological fields, not only nationally but also internationally. The IT sector in Germany is growing through a remarkable evolution process. Companies in the IT sector are continuously innovating products to attract customers in the face of digitalization.

This led to several people entering the IT industry, as there are big open doors for newcomers to kick start their careers and experienced people are looking for better opportunities. To understand the Information Technology market in Germany, a survey was conducted for three consecutive years from 2018 to 2020 to assess the importance of skill sets for IT specialists in relation to experience, position, and other key factors.

**Dataset selected**

                       The dataset acquired from “Kaggle” contains the salary information of IT professionals and the data was collected anonymously through Microsoft forms for 3 years starting from 2018. During the inauguration year 2018, The questions were straightforward and the count kept increasing over the next couple of years including some of the predominant questions like skill set, vacation days per year and even covered the epic corona pandemic.

Dataset source - <https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>

**Background**

                       There are positions in the IT world that are high in demand and where there is a shortage of qualified workers. In addition to indicating mid-to senior-level positions, these professions are also associated with higher incomes. A lower salary is generally associated with entry-level positions. These include positions such as help desk analyst, IT technician, and IT associate. One will see your salary increase as you gain more experience and specialize in areas like systems administration or cloud computing.

By acquiring in-demand skills through certifications or other means, salaries have increased. This is also true of furthering your education. A survey by Global Knowledge found that 12 percent of those who received a raise in 2020 attributed it to new skills, whether they achieved certifications or not. The average salary increase of IT professionals who got new certifications was $13,000.

**Motivation**

                       The salary analysis serves a variety of goals, ranging from wage determination to market competitiveness to internal salary equity problems as they occur. With changes in employment markets and supply and demand, the statistics in a wage analysis will fluctuate throughout time, shifting up and down based on need and relevance.

Salary surveys conducted over a few years can assist human resource professionals, Job consulting firms and Job seekers in determining new roles, compensation increases, volatility, trends, or jobs on the decrease.

**Insights**

                       It was unheard of in the past for tools to help analyze which technologies are in demand, giving employees who are looking to switch companies the opportunity to earn a better salary. Consultancies, recruitment departments, and startup companies would need to spend a great deal of time finding out about emerging technologies and the packages they can offer without these tools and data.

The IT employees can evaluate the information obtained from the analysis and comprehend the technologies that are in demand, as well as the packages being offered by companies for each technology without utilizing the services of consultants.

**Problem statement**

                       The objective is to estimate the expected salary level of each job position and most promising skill set every year based on the survey results for Germany from 2018 to 2020, considering multiple factors like years of experience, work locations, and contract types. The big tech companies hire or retain many employees based on industry requirements. A person might find it confusing to determine the right career path due to the wide range of career opportunities in the IT sector in Germany.

Salary planning is one of the most important sub-disciplines of human resources and plays an essential role in attracting, motivating, and retaining talents. One important part of this planning is salary benchmarking, which aims to determine the market pay scales of employees with respect to different job roles. This analysis of survey results will help the HR team in the big companies, Consulting firms and Individual job seekers for a better understanding of the market.

**Project Proposal**

                       We will be curating a descriptive analysis with the dataset, answering most of the hypothetical questions that pop up when seeking a job in a corporate firm and providing insights to the job consulting firms about the Information Technology industry as well as Human resource professionals working in the same industry. In this manner, they can assess their candidates based on their level of work experience and waive off other required skill sets helping them to land their dream job.

We will examine the highest paid salaries, the most popular programming languages, and the seniority and their pay based on these languages here. A key goal of this study is to estimate the competitive value of an IT skill set based on the years of experience, place, salary, etc.

**Target Audience**

                       The Job consulting firms that are based in Germany and Human resource professionals n will be our target, giving them the major insights about the Cost to company (CTC) discussions considering various factors of an employee along with the top skills that are required to get a job from a lucrative organization.

Getting a decent pay job nowadays is a hustle when you are new to the market and the same is applicable for experienced employees as well. The consulting firms should be all ears to the market, knowing what they should do to get a job seeker inside a corporate firm. This project will help the firm to seek a naive employee who is looking for a job, consult him and advise the current market strategy to end up getting them a high pay job.

**Scope**

Start-ups and Human Resource professionals

                       In this new era of technology where every individual is aspiring to start their own company. Based on this information, the company will be able to determine the employee's career path and their position within the company, as well as the type and area of business the company engages in.

 Job Consultancy firms

                       Most consultancies analyze these data to find the most sought-after technologies with highly competitive salaries and then recommend training people for those technologies.

**Analysis questions**

* Which age group has the most responses?
* Which city has the greatest number of respondents?
* What is the median salary of Male, females and Other sexualities?
* What are the top 5 skill sets of respondents?
* What language do most of the respondents speak at work?
* What is the total percentage of people who lost their jobs due to the Corona outbreak?
* What are the top 5 high-paying business sectors?
* What is the most common position of the respondents?
* How many respondents are Freelancers?

**EXPLORATORY DATA ANALYSIS(EDA)**

Logo

Description automatically generated

**KNOW YOUR WORTH!**

**Information Technology Salary Survey for Europe region**

**Dataset description**

                       An anonymous salary survey has been conducted annually since the year 2018 among European IT specialists with a stronger focus on Germany. The data has been made publicly available by the authors. The dataset contains rich information about the salary patterns among the IT professionals in the EU region and offers some great insights about the top skills and other key factors to sustain their position in the market. A study of IT individuals over a 3-year period depicted a gradual rise in participation by about a quarter percent.

                       The dataset contains 3 CSV files for the consecutive years 2018, 2019 and 2020. There are 23 features and 3K observations each representing an IT professional with different backgrounds and skill sets. The predominant feature of the dataset that discusses the corona outbreak gives an idea of lay off and recession that happened during the pandemic.

**Column description and data type**

|  |  |  |
| --- | --- | --- |
| COLUMN NAME | DATA TYPE | DESCRIPTION |
| Timestamp | Date | Date of entry |
| Age | Numerical | Age of the respondent |
| Gender | Categorical | Gender of the respondent |
| City | Categorical | City of the respondent |
| Position | Categorical | Role at the working firm |
| Total years of experience | Numerical | Overall total experience |
| Years of experience in Germany | Numerical | Work experience in Germany |
| Seniority level | Categorical | Designation at the working firm |
| Your main technology/ programming language | Categorical | Primary skill set |
| Other technologies/programming languages you use often | Categorical | Secondary skill set |
| Yearly Brutto salary (without bonus and stocks) in EUR | Numerical | Annual salary excluding Bonus and Stocks |
| Yearly bonus + stocks in EUR | Numerical | Annual Bonus and Stocks |
| Annual Brutto salary (without bonus and stocks) one year ago. Only answer if staying in the same country | Numerical | Previous annual salary excluding Bonus and Stocks |
| Annual bonus+stocks one year ago. Only answer if staying in same country | Numerical | Previous annual Bonus and Stocks |
| Number of vacation days | Numerical | Total number of vacation days per year |
| Employment status | Categorical | Employment type of the respondent |
| Ð¡ontract duration | Categorical | Contract duration of the respondent |
| Main language at work | Categorical | Communication language at the workplace |
| Company size | Numerical | Size of the company |
| Company type | Categorical | Business unit of the organization |
| Have you lost your job due to the coronavirus outbreak? | Categorical | Did you lose your job due to the Corona pandemic? |
| Have you been forced to have a shorter working week (Kurzarbeit)? If yes, how many hours per week | Numerical | Have you been given the work hours you suppose to work? |
| Have you received additional monetary support from your employer due to Work from Home? If yes, how much in 2020 in EUR | Numerical | Did your firm help with work from allowance? |

**Incomplete, missing or Invalid records**

                       The dataset contains missing values, and the below table encompasses the number of missing values per column from the dataset.

Graphical user interface, text, application, email

Description automatically generated

**Data segmentation**

                       The data can be segmented with respect to features like Contract duration, Employment status and Seniority level of the Organization. The imperative way to segment the dataset is by splitting the observations corresponding to the Year of Survey taken and Age group.

**Data imbalance**

                       The data is slightly imbalanced as we have almost 84% of the respondents are Male and 43% of the survey responses were from Berlin city from Germany. Since this survey was conducted throughout Germany for the purpose of recognizing salary patterns of IT professionals, this data imbalance is legitimate.

**Data correlation**

Chart

Description automatically generated

                       The current gross salary and the salary one year ago of the respondents are positively correlated other than the same features of the dataset and a negative correlation can be spotted between other features of the dataset.

**Data collection**

This survey data was collected through Microsoft forms for three consecutive years 2018, 2019 and 2020 by “Viktor Shcherban’’ and “Sergey Vasilyev”. This anonymous survey made respondents answer 23 questions regarding their corporate life. As we know we had to go through a pandemic. During the initial months of 2020 which eventually snatched many respondent jobs since the market was completely down and data contains this as a question which is addressed. We spotted this dataset from “Kaggle”.

Dataset source - <https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>

Survey link - <https://docs.google.com/forms/d/e/1FAIpQLSdPDpjEN98tazCLOQ7xxgK84DZeanC8wI_akPyKOeW3HwBhuA/viewform>

Here are a few anonymous questions from the survey.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Outliers**

* Age outlier

Chart, scatter chart

Description automatically generated

                       There is a right skewed outlier when it comes to Age feature since we had minimal responses from people above 50 years of age. Since most of the employees begin their career at the age of 20, There would be a left-skewed outlier that is redundant when it comes to real-world data of IT professionals.

* Total experience of the respondent

Chart, line chart

Description automatically generated

                       Most of the respondents who took the IT salary survey fall under less than 20 years of experience. We can recognize a right-skewed outlier from this chart as people who work more than 20 years are not ready to take up the survey. This can be neglected since we majority of the respondents lie in the first set of categories.

**Inclusion Criteria**

                       Geographically, the dataset is based on Germany dated from 2018-2020. We are planning to add few unique columns like Survery\_id, Total\_Annual\_salary (Including yearly+bonus+stocks), Previous\_Year\_Total\_Salary and make use of these columns segregating fact and dimension tables. The inclusion criteria for the dataset are

* Gender
* Age
* Position
* Primary skill set
* Yearly Gross salary

**Schema diagram**

Diagram

Description automatically generated

**Preliminary Visualizations**

* IT professionals participation in the survey

A picture containing diagram

Description automatically generated

The number of people participating in the Information technology salary survey every year seems to be gradually increasing. The number of respondents every year surge by quarter.

* Age of respondents Vs Number of responses (2020)

Chart, bar chart, histogram

Description automatically generated

The below chart conveys Age group between 24-38 are the active participants of the survey. The respondents aged 30 contributed 9% of the total responses which being the top contributor.

* The proportion of Contract type of the respondents (2020)

Chart

Description automatically generated

From the above graph, most of the respondents are permanent workers. 95% of the IT professionals who took the surveys have an unlimited contract with their firm and 5% of the respondents are Contractors.

* The frequency of vacation days per year (2020)

Text

Description automatically generated

Most of the employees get 30 paid annual vacation days per year contributing about 7 times those who get 26 days per year from the entire responses of the 2020 survey.

* Company size Vs Number of responses (2019)

Chart

Description automatically generated

There are very few respondents who work in start-ups that fall under company size less than 10 and almost half of the respondents work in a reputed firm. Twenty-five percent of respondents work for companies with 50 employees.

* Gender proportion of the respondents

Chart, pie chart

Description automatically generated

More than the 3rd quartile of the responses is from Male and 16% of the total respondents are female. Diverse people seem to be very less in number contributing less than 0.5% of the total responses.