**CAPSTONE PROJECT REPORT**

**Global Software Engineer Salary Trends: An In-Depth Country Analysis**

**Submitted by**

**V.Gnaneswar Reddy (192224182)**

**M.Karthik Reddy(192224185)**

## **Guided by**

V.Saranya

Junior Research Fellow

Department of Neural Networks

Department of Computer Science and Engineering,

Saveetha School of Engineering, SIMATS Thandalam, Chennai

March – 2024

**PROBLEM STATEMENT:**

The software engineering field is expanding globally, but salary variations across regions, skills, and experience levels create challenges for both job-seekers and employers. Without consistent, detailed data, it’s difficult to set fair salary benchmarks or identify valuable skills by region. Existing salary information is often fragmented, limited to average figures, and lacks depth on key variables like specific skills or detailed experience tiers. This project aims to address these gaps by analyzing software engineer salary trends across multiple countries and skill sets, using Python for data processing and Matplotlib and Power BI for dynamic, interactive visualizations. By presenting comprehensive, accessible insights into how salaries vary based on country, expertise, and skill demand, this study will support software engineers in career planning and help employers set competitive, fair compensation structures globally.

### **EXISTING SYSTEM OF APPROACH:**

1. **Data Collection and Processing**: Conventional approaches rely on sources like Glassdoor, Payscale, and LinkedIn. However, these sources can be limited by inconsistencies and incomplete data, complicating trend analysis.
2. **Basic Statistical Techniques**: Traditional methods, including simple regression and trend extrapolation, provide a broad overview but lack the precision to capture complex regional and economic interactions in salary trends.
3. **Static Visualizations**: Salary data is often presented in static charts and tables, limiting user engagement and making it challenging for non-expert stakeholders to explore nuances in the data.

### **PROPOSED SYSTEM OF APPROACH:**

1. **Data Automation and Processing with Python**: Automated data gathering and preprocessing using Python libraries (Pandas, NumPy) ensure consistent, reliable data, with comprehensive handling of anomalies and missing values.
2. **Advanced Statistical Techniques**: Time series and regression models improve accuracy in predicting salary trends by analyzing the influence of specific economic factors and tech skill demands on global salaries.
3. **Dynamic Visualizations with Matplotlib and Power BI**: Matplotlib enables precise plotting of trend lines and salary distributions, while Power BI creates interactive dashboards, allowing users to explore data by region, role, and skill set.

#### **Advantages of the Proposed Approach:**

1. **Improved Data Quality**: Automation and rigorous preprocessing ensure high-quality, up-to-date data for analysis.
2. **Enhanced Forecasting Precision**: Advanced statistical methods provide more nuanced predictions, accounting for complex regional and skill-based factors.

**Increased User Engagement**: Interactive visualizations make data more accessible, allowing stakeholders to explore and interpret salar

**DATASET ANALYSIS :**

The primary dataset used in this project includes the following key fields:

**Job Title:** Represents the specific software engineering role (e.g., Front-End Developer, Data Scientist etc).

**Skills:** Divided into "Key Skill 1" and "Key Skill 2," these fields capture the technical competencies most relevant to each job role (e.g., JavaScript, Python, Machine Learning etc).

**Country:** Indicates the country in which the salary data is reported, covering a diverse range of regions for comprehensive global analysis.

**Average Salary (USD):** Provides the average salary for each job role within each country, allowing for cross-country comparisons.

**Experience Level:** Categorized as Entry, Junior, and Senior, indicating the professional level of the software engineer, which plays a crucial role in salary determination.

**Visualization with Matplotlib and Power BI:**

Data visualization is a crucial component of this analysis, enabling stakeholders to interpret complex salary patterns in a visually intuitive manner. Python's Matplotlib library is used to create foundational graphs, such as country-specific salary distributions, skill-demand correlations, and salary trends by experience level. These static visualizations provide insights into the broad trends and initial patterns within the dataset.

To provide interactive, dynamic visualization capabilities, Power BI is employed. Power BI dashboards facilitate a deeper exploration of the data, allowing users to filter results by country, job title, or skillset to examine specific trends. Features like heatmaps for salary variation across countries and time-series plots for salary growth within individual regions make the data accessible and engaging. Through these combined visualization tools, the project enhances user engagement and delivers a powerful platform for exploring global software engineer salary trends.

**ENVIRONMENTAL SETUP:**

**1. Install Python and Necessary Libraries:**

* Python: Download from official website.
* Jupyter Notebook: pip install notebook.

2. **Install Required Libraries:NumPy:**

* pip install numpy.
* Pandas: pip install pandas.
* Matplotlib: pip install matplotlib.

**3. Set Up Working Directory:**

* Create a directory for project files.

**4. Obtain Population Data:**

* Download from sources like United Nations or World Bank.

**Power BI Environment Setup:**

**1. Install Power BI Desktop:**

* Download from official Power BI website.

**2. Prepare the Data:**

* Obtain data in CSV or Excel format from reliable sources.

**3. Import Data into Power BI:**

* Use “Get Data” feature to import population data.
* Clean and transform data using Power Query Editor.

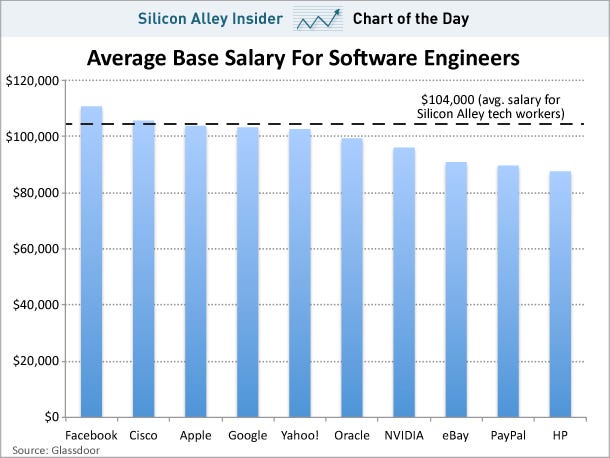
**4. Create Visualizations:**

* Use drag-and-drop interface for charts, graphs, and maps.
* Utilize built-in mapping tools for geographic visualization.

**5. Publish and Share Dashboards:**

* Publish reports to Power BI service for sharing and collaboration

**DATA FLOW DIAGRAM :**



**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv('salary\_data.csv')

print(df.head())

df.dropna(subset=['Entry Salary', 'Junior Salary', 'Senior Salary', 'Country'], inplace=True)

df\_melted = pd.melt(df, id\_vars=["Job Title", "Country", "Skills"],

value\_vars=["Entry Salary", "Junior Salary", "Senior Salary"],

var\_name="Experience Level", value\_name="Salary")

df\_melted['Experience Level'] = pd.Categorical(df\_melted['Experience Level'],

categories=["Entry Salary", "Junior Salary", "Senior Salary"],

ordered=True)

plt.figure(figsize=(12, 8))

sns.barplot(x='Salary', y='Country', hue='Job Title', data=df\_melted, palette='viridis')

plt.title('Global Salary Trends by Job Title and Country')

plt.xlabel('Salary (USD)')

plt.ylabel('Country')

plt.show()

plt.figure(figsize=(10, 6))

sns.barplot(x='Experience Level', y='Salary', hue='Job Title', data=df\_melted, palette='coolwarm')

plt.title('Salary Trends by Experience Level and Job Title')

plt.xlabel('Experience Level')

plt.ylabel('Salary (USD)')

plt.show()

plt.figure(figsize=(12, 8))

sns.boxplot(x='Experience Level', y='Salary', data=df\_melted, palette='muted')

plt.title('Salary Distribution by Experience Level')

plt.xlabel('Experience Level')

plt.ylabel('Salary (USD)')

plt.show()

salary\_columns = ['Entry Salary', 'Junior Salary', 'Senior Salary']

correlation\_matrix = df[salary\_columns].corr()

plt.figure(figsize=(8, 6))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)

plt.title('Correlation Heatmap of Salary Levels')

plt.show()

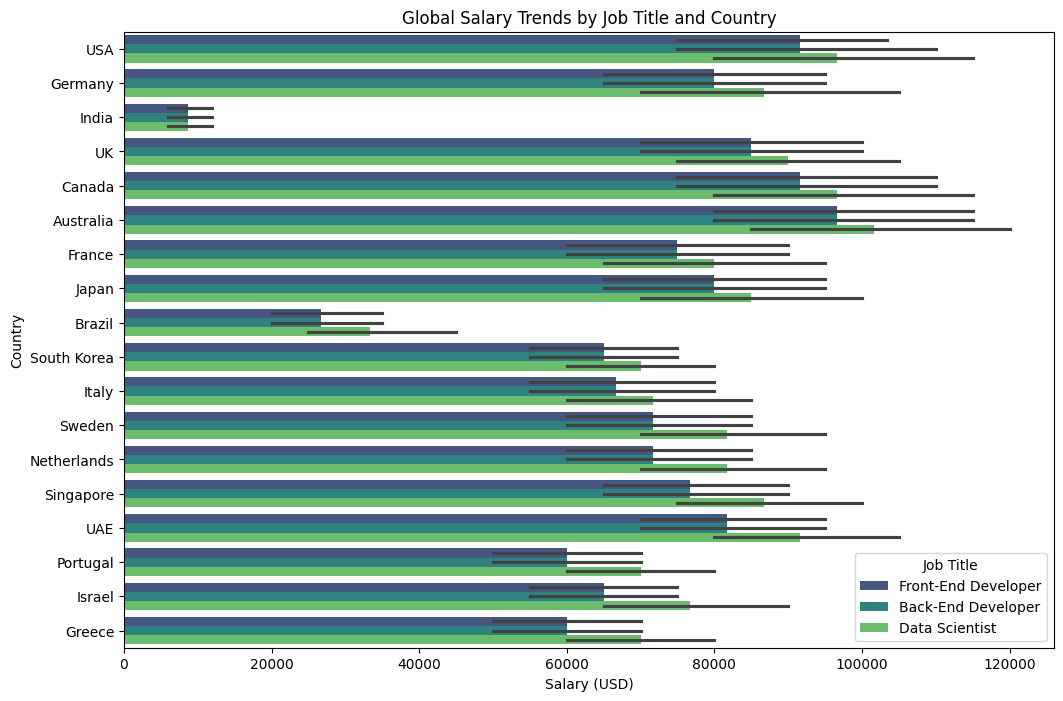
**RESULT ANALYSIS :**

This analysis of global software engineer salary trends reveals the nuanced interplay between geography, skill set, experience, and broader economic forces in shaping compensation. Salaries in the tech sector are not only dictated by market demand but are also sensitive to local and global economic indicators, workforce availability, and rapid shifts in skill requirements driven by technological evolution. Highly specialized skills, such as machine learning or cloud computing, command higher wages, particularly in regions where there is a scarcity of expertise or intense competition for talent. Additionally, entry-level roles tend to reflect regional economic standards, whereas senior positions align more closely with global pay scales, often adjusting to attract top talent.

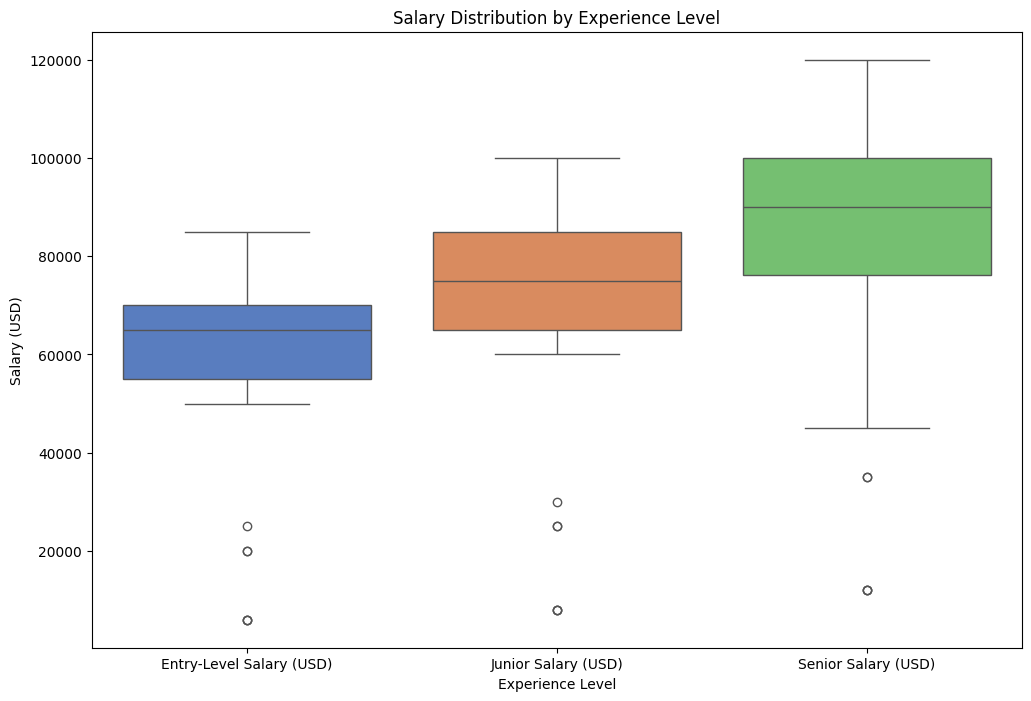
From this study, we can observe how the rise of remote work and digital collaboration has begun to reshape regional wage structures, potentially leading to greater salary equilibrium across borders. For professionals, understanding these trends provides a data-driven approach to career planning, enabling informed decisions on skill development and geographic mobility. For employers, these insights support strategic workforce planning and optimized resource allocation, fostering competitiveness in a global talent market.

By merging skill-based analysis with geographic and economic factors, this study contributes to a comprehensive understanding of software engineering compensation trends, offering valuable insights for academia, industry, and policymakers alike. The findings underscore the critical role of data analytics in adapting to and planning for the ever-changing landscape of global tech employment.

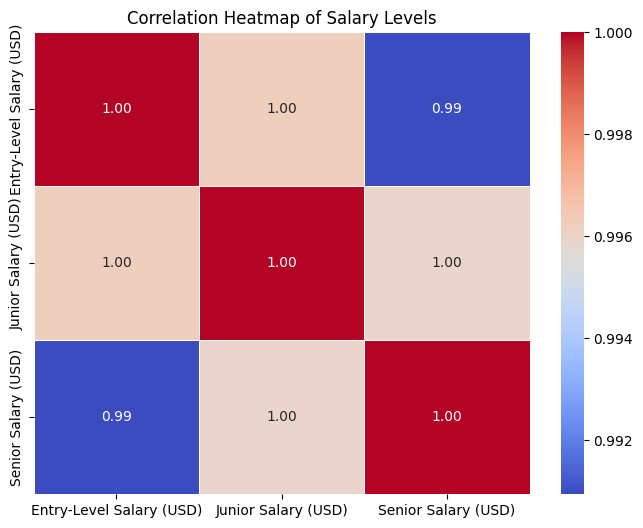
**OUTPUT SAMPLES :**

****

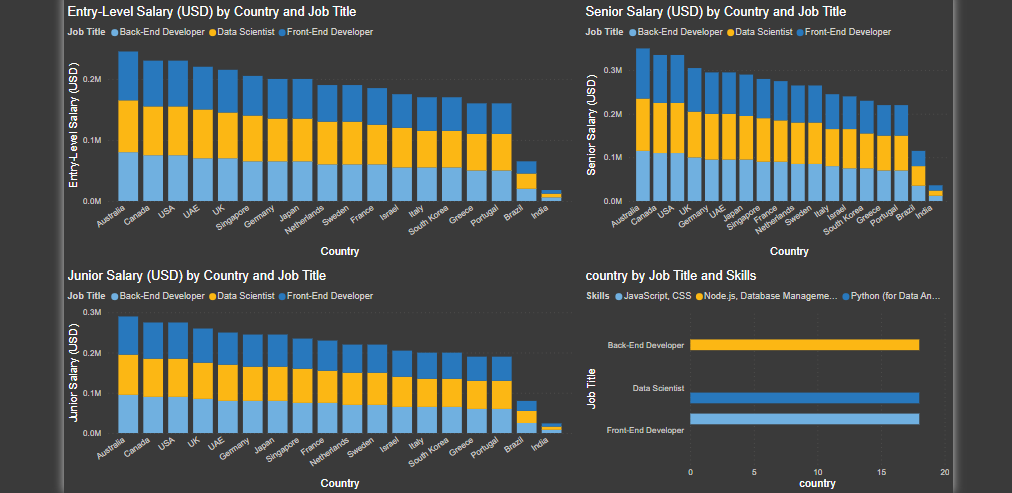
**FIG1: Global salary trends by job title**

****

**FIG2: Boxplot of salary distribution by experience level**

****

**FIG3: Heatmap of salary levels**

****

**FIG4: PowerBI dashboard**