

Exploring Bachelor's Degree Employment Trends in the USA

In today's world, the value of a college degree is often measured by the job opportunities it opens. As students put in a lot of effort and money to earn bachelor's degrees, it's important to know what kind of jobs and salaries they can expect after graduation. This project uses a detailed set of data from Kaggle to look into how well bachelor's degree holders are doing in finding jobs and earning a good salary. We aim to show how different college majors and schools affect students' chances in the job market. We also aim in the project report to discuss the python-based tools we used and how we incorporated cloud in our project.

This project analysis is not just helpful for students thinking about their future careers. It also helps teachers design better courses and government leaders make smarter education policies. By looking at what helps students get good jobs and how these factors differ by major and school, our study provides important information that can help improve education and career planning.

We used tools and methods for our analysis, mainly using python, a programming language, and its libraries like pandas, matplotlib, and seaborn. These tools help us manage data, create graphs, and understand complex information. We also used the Google Colab platform that has the jupyter notebook installed as a way to implement cloud into our project. We also explore how well students do based on their grades, how much experience they have, and where they went to school.

The main areas we focus on include:

- Looking at how much graduates earn depending on their major and college.
- Seeing if there's a link between students' grades and how much they earn, to show the importance of doing well in school.
- Examining how work experience gained during college, like internships, affects job chances and starting salaries.
- Comparing job placement rates among different colleges to see which ones are doing a better job at helping their students succeed.

We hope the information we find will help many people. Students will be able to make better choices about where to go to college and what to study. Schools can use this information to improve their programs, ensuring they meet the needs of the job market.

In the job market, the value of a college degree extends far beyond academic achievement—it serves as a gateway to diverse career opportunities and financial stability. Recognizing the

significant investment of time and resources made by students pursuing bachelor's degrees, it becomes important to check the correlation between educational choices and subsequent employment outcomes. This project goes into a dataset sourced from kaggle, which aims to find the trajectories of bachelor's degree holders in their professional journeys.

Beyond serving the immediate needs of prospective students navigating their academic futures, this analysis holds bigger implications for educational stakeholders and policymakers alike. By using different methodologies and python-based tools such as pandas, matplotlib, and seaborn, we go into the correlations between majors, institutions, and employment prospects. Also including Google Colab, our investigation spans across multiple outlooks:

First, we look at how much money graduates make based on what they studied and where they went to school. Then, we see if getting good grades in college leads to higher salaries after graduation. We also check out how doing internships or other hands-on learning experiences during college affects getting a job and how much money you start with. Finally, we compare how well different colleges help their graduates find jobs, so you can see which ones might be better at preparing students for the workforce.

By putting all this together, our study helps people choose the right education and career path. It also gives schools useful information to improve their programs and better prepare students for jobs. In the end, we want to make sure students have what they need to succeed in their careers, even as things change over time.

The project's design architecture was created to manage the analysis of employment outcomes for bachelor's degree graduates. This architecture has several key design decisions that are aimed at optimizing data preprocessing, cloud computing, analysis, and visualization processes.

Firstly, data preprocessing is prioritized to ensure the dataset's readiness for analysis. This involves converting the 'placement_status' column to numeric values and handling non-numeric salary data gracefully, ensuring a smooth and error-free analysis process. By addressing data inconsistencies upfront, we lay a solid foundation for meaningful insights.

Secondly, a modular approach is adopted to break down the analysis into specific sections. This modular structure enhances clarity and ease of understanding, which allows stakeholders to navigate through different aspects of the analysis. Specific sections include overall placement rates, salary distribution, factors influencing placement success, and comparison among universities.

Lastly, visualization plays a crucial role in conveying insights derived from the analysis. Bar plots and scatter plots with trend lines are chosen as the primary visualization tools due to their

effectiveness in representing data trends and patterns. These visualizations provide stakeholders with a clear and intuitive understanding of employment outcomes, which enables informed decision-making.

The implementation process follows a systematic approach, starting with data loading and culminating in analysis and visualization stages. We wanted to make sure that our implementation was correct so that it could be compiled in the jupyter notebook with Google Colab.

Data loading involves importing the dataset into a Pandas DataFrame, ensuring all relevant fields such as 'placement_status', 'salary', field of study, university, GPA, and years of experience are included. These fields will then be converted into the bar and scatter plot graphs for data analysis.

Data preprocessing is then carried out to convert the 'placement_status' column to numeric values and handle non-numeric salary data appropriately. This step is essential for preparing the dataset for analysis and ensuring data consistency. Below, we can see this implemented in python code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
df = pd.read_csv('job_placement.csv')

# Convert 'placement_status' to numeric for analysis (1 for Placed, 0 for Not Placed)
df['placement_numeric'] = df['placement_status'].apply(lambda x: 1 if x == 'Placed' else 0)

# Filter out non-placed candidates for salary analysis
placed_df = df[df['placement_status'] == 'Placed']

# Ensure 'salary' is numeric and handle non-numeric gracefully
df['salary'] = pd.to_numeric(df['salary'], errors='coerce')

# Salary Distribution Among Graduates Based on Field of Study and University
salary_distribution = placed_df.groupby(['stream', 'college_name'])['salary'].mean()
plt.figure(figsize=(14, 7))
salary_distribution.plot(kind='bar')
```

```

plt.title('Average Salary by Field of Study and University')
plt.ylabel('Average Salary')
plt.xlabel('Field of Study')
plt.xticks(rotation=45, ha='right')
plt.legend(title='University', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()

# Scatter plot for GPA vs. Salary with trend line
plt.figure(figsize=(10, 6))
sns.regplot(x='gpa', y='salary', data=placed_df, scatter_kws={'alpha':0.5}, li
plt.title('GPA vs. Salary for Placed Graduates')
plt.xlabel('GPA')
plt.ylabel('Salary')
plt.grid(True)
plt.show()

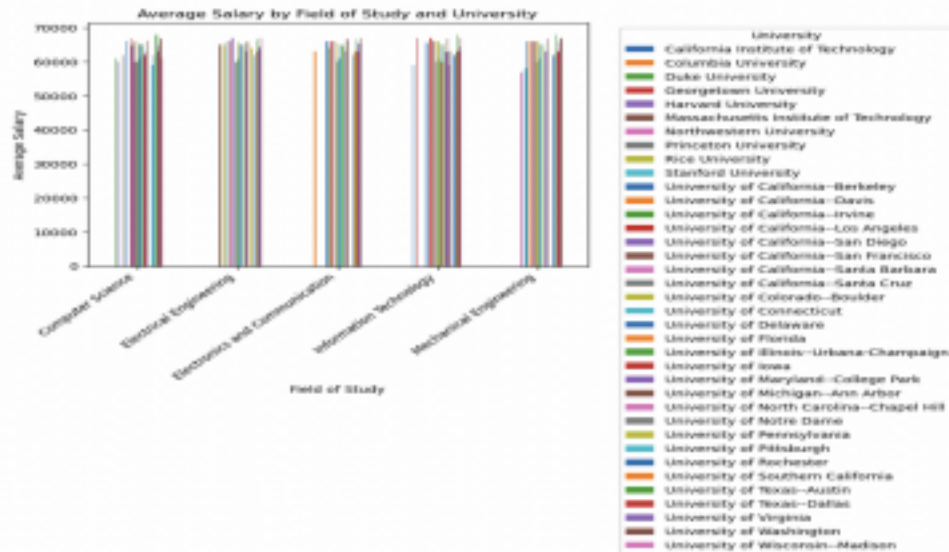
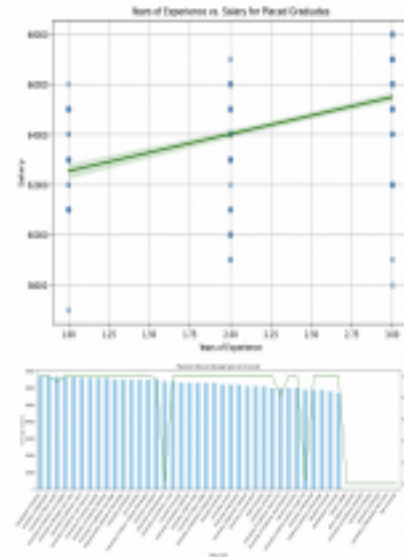
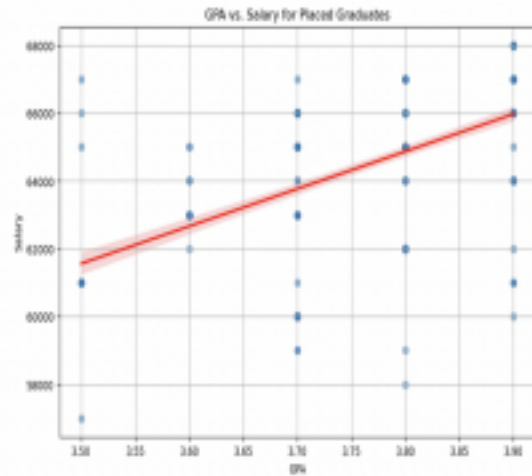
# Scatter plot for Years of Experience vs. Salary with trend line
plt.figure(figsize=(10, 6))
sns.regplot(x='years_of_experience', y='salary', data=placed_df, scatter_kws={
plt.title('Years of Experience vs. Salary for Placed Graduates')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.grid(True)
plt.show()

# Compare Placement Rates and Average Salaries Among Universities
university_analysis = df.groupby('college_name').agg(
    placement_rate=('placement_numeric', 'mean'),
    average_salary=('salary', lambda x: x[x > 0].mean())
).sort_values(by='average_salary', ascending=False)

fig, ax1 = plt.subplots(figsize=(14, 7))
university_analysis['average_salary'].plot(kind='bar', color='skyblue', ax=ax1
ax2 = ax1.twinx()
university_analysis['placement_rate'].plot(kind='line', color='green', ax=ax2)
ax1.set_title('Placement Rate and Average Salary by University')
ax1.set_ylabel('Average Salary')
ax2.set_ylabel('Placement Rate')
ax1.set_xticklabels(university_analysis.index, rotation=45, ha='right')
plt.tight_layout()
plt.show()

```

Analysis and visualization stages involve conducting various analyses, such as examining salary distribution by field of study and university, analyzing the relationship between GPA and salary, exploring the impact of years of experience on salary, and comparing placement rates and salaries among universities. Each analysis is accompanied by visual plots, including bar plots, scatter plots, and combined bar and line plots, enhancing the interpretability of the results. Below, we can see the graphs that show that data we have discussed previously:



Overall, our implementation process adheres to the established design architecture, ensuring a systematic and rigorous approach to analyzing employment outcomes for bachelor's degree graduates. We have also established our connection for cloud purposes as we utilized Google Colab.

The analysis conducted in this project yields insightful findings regarding the employment outcomes of Bachelor's degree graduates in the USA. By examining various factors such as salary distribution, GPA vs. salary relationships, and university comparisons, the project provides a comprehensive understanding of the dynamics shaping employment trends. One important insight uncovered through the analysis is the impact of academic performance on employment

outcomes.

The exploration of the relationship between GPA and salary reveals a positive correlation, highlighting the importance of academic achievement in securing various job opportunities. This finding emphasizes the value of academic excellence in enhancing students' competitiveness in the job market and clarifies the importance of maintaining high academic standards in educational institutions.

Furthermore, the analysis delves into the influence of work experience, particularly internships or hands-on learning experiences, on job placement and starting salaries. By examining the relationship between years of experience and salary, the project elucidates the role of practical training in augmenting graduates' employability and earning potential. The findings suggest that students who engage in internships or gain relevant work experience during college are more likely to secure employment and command higher salaries upon graduation. This highlights the significance of experiential learning opportunities in bridging the gap between academic knowledge and real-world application, thereby equipping students with valuable skills and competencies sought after by employers.

Moreover, the project compares job placement rates and salaries among different universities, shedding light on the varying levels of success in preparing graduates for the workforce. The analysis reveals disparities in employment outcomes across institutions, with some universities demonstrating higher placement rates and average salaries than others. This underscores the importance of institutional factors, such as curriculum design, career services, and industry partnerships, in shaping students' career trajectories. By identifying universities that excel in preparing graduates for successful employment, the project offers valuable insights for prospective students, educators, and policymakers in evaluating and improving higher education programs.

Overall, the measurement and experiment results provide valuable insights into Bachelor's degree employment trends in the USA, elucidating the multifaceted factors influencing graduates' career outcomes. The findings underscore the importance of academic performance, work experience, and institutional support in shaping students' employability and earning potential. By using different methodologies, data analysis, and cloud computing the project contributes to informed decision-making in education and career planning, empowers students to make strategic choices for their future success.

In this discussion portion for our project, we will discuss how we aimed to analyze the employment outcomes of Bachelor's degree graduates, focusing on trends in job placement, factors influencing placement success, and salary distribution across different fields of study and universities. Our approach was the utilization of cloud computing technologies along with Python-based tools to enhance the efficiency and scalability of our analysis.

Initially, our plan involved implementing cloud computing through Azure and utilizing Apache Hadoop to process and analyze the dataset sourced from Kaggle. However, we encountered limitations related to Azure's free credit limit, which constrained our ability to build and maintain virtual machines for data processing. We recognize the importance of cloud computing in facilitating data analysis from kaggle, so we searched for an alternative solution that would allow us to use cloud resources effectively.

Therefore, we transitioned our project to Google Colab, which is a cloud-based platform that provides free access to computational resources and integrates seamlessly with Python, Pandas, and other essential libraries for data analysis. By using the Google Colab platform, we were able to overcome the constraints that were imposed by Azure's credit limit while still using the benefits of cloud computing for our project.

One of the key advantages of using the Google Colab platform was its ease of use and accessibility, allowing us to collaborate on the analysis in real-time and access computing resources without the need for extensive setup or configuration. Additionally, Google Colab's integration with Google Drive enabled seamless data sharing and storage, facilitating efficient data management throughout the project.

Our decision to continue emphasizing cloud computing aspects in our project stems from several key factors such as:

First is scalability because cloud computing offers unparalleled scalability, enabling us to dynamically adjust computing resources as our project evolves and our dataset grows. Second is resource accessibility because cloud-based platforms like Google Colab provide access to a wide range of computational resources without the need for costly infrastructure investments, democratizing access to advanced computing capabilities.

We also have collaboration and accessibility so that the cloud computing environments facilitate collaboration and accessibility, which fosters greater teamwork, knowledge sharing, and productivity among project team members. Then, we have innovation and experimentation because cloud computing environments provide opportunities for innovation and experimentation, allowing us to explore new technologies, tools, and methodologies in a risk-free environment. The significance of python code in Google Colab platform with the jupyter notebook was also another key factor.

We also created two presentations for the class to showcase our project and update our findings for the dataset on kaggle. Our goal was to essentially analyze the data by using python and also use the Google Colab platform as a means to upload it to the cloud. It was important for us to

present our findings so that we could receive feedback from students so that we can continue to improve our method and incorporate ways to add cloud computing to our project.

The use of Python code in Google Colab with Jupyter Notebook offers several significant advantages such as:

Virtual machine equivalent was what we were looking for to implement a form of a cloud system so we went for Google Colab because it provides users with their version of a virtual machine (VM) environment, which closely resembles a traditional VM setup and offering computational resources and an isolated workspace for code execution and analysis. Also, hands-on learning because working with python code in Google Colab provides a hands-on learning experience in cloud computing, enabling team members to gain practical insights into cloud computing concepts and best practices.

Another advantage is real-world relevance because python code executed in Google Colab mirrors real-world scenarios commonly encountered in data analysis and machine learning projects, enhancing team members' proficiency and marketability in professional settings. Finally, integration with cloud services because Google Colab integrates with various cloud services and APIs, enabling access to additional resources and functionalities, such as cloud-based databases and machine learning services, to augment our analysis and drive innovation.

In conclusion, by incorporating cloud computing aspects and utilizing Python code in Google Colab with jupyter notebook, we not only overcome the challenges posed by resource limitations but also gain valuable insights and skills that are directly applicable to real-world scenarios. Moving forward, we are committed to further exploring and integrating cloud computing techniques to enhance the depth and sophistication of our analysis, so we can deliver impactful results in our examination of employment outcomes for bachelor's degree graduates.