**BAIS:3200 Final Project**

**Introduction**

The modern job market is increasingly shaped by data analysis, which influences hiring trends and the skills employers seek. In this project, we perform descriptive analysis on a dataset of job listings spanning various industries, locations, experience levels, and salary ranges. By examining these patterns, we aim to generate insights that support AI-driven job recommendation systems. This analysis will be valuable for data scientists, HR professionals, job seekers, and researchers who are interested in using data-driven methods to understand employment trends and improve career decision-making.

**Data**

This project uses a cleaned and normalized subset of a job recommendation dataset originally sourced from Kaggle[[1]](#footnote-1). The full dataset contains over 50,000 job listings compiled from company websites, recruitment platforms, and online job portals. To ensure that our analysis was manageable and fast in Oracle APEX, we randomly sampled 5,000 job postings using R. This made our analysis much more digestible and meaningful. These job listings, as referenced above, span a variety of different industries, experience levels, locations, and salary ranges. Table 1 provides a detailed outline of each field, data type, and a description of its purpose in the dataset.

*Table 1 Data Dictionary*A black and white text

AI-generated content may be incorrect.

In our ER diagram (Figure 2), we modeled the dataset using two entities: COMPANY and JOB. The COMPANY entity represented organizations posting job listings, uniquely identified by a CompanyID. It included attributes such as CompanyName and Location. This entity served as the foundation for the data model, as all job postings in the dataset are associated with a specific company.

A line with black dots and words

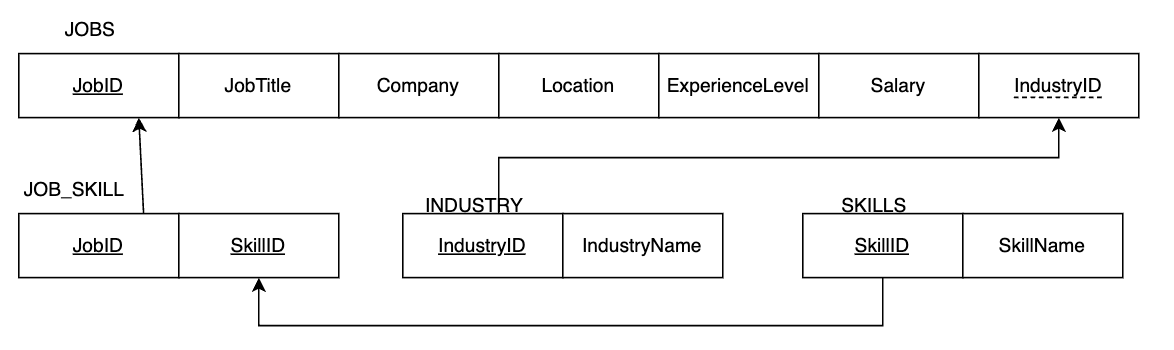
AI-generated content may be incorrect., Picture

*Figure 2 Entity relationship diagram*

The JOB entity represented individual job listings and was modeled as a weak entity, meaning it depended on the existence of a parent company. Each job record included required attributes such as Experience Level, Salary, Industry, and Skills. The Skills attribute was multivalued, representing a set of technical or soft skills required for the job. In the ER diagram, the COMPANY and JOB entities were connected through a one-to-many relationship: a single company could post multiple jobs, but each job must be associated with exactly one company.

This conceptual model reflected the real-world relationship between employers and job openings. It also highlighted data constraints, such as ensuring every job is linked to an existing company and that each company has at least one active listing. The diagram also acknowledged the complexity introduced by the multivalued Skills attribute, which informed our later design choices.

While our original conceptual model included a distinct COMPANY entity, we revised the implementation during normalization. In the final schema, Company is treated as a descriptive field within the JOBS table, and the multivalued Skills attribute is separated into a SKILL table and an associative JOB\_SKILL table. This transformation helped us support better data integrity, while also enabling us to have more flexible querying in Oracle APEX.



*Figure 3 Graphical relational schema*

To support database normalization and improve data integrity, we implemented a revised relational schema based on our conceptual model. The final design consists of four normalized tables: JOBS, INDUSTRY, SKILL, and JOB\_SKILL. These tables were structured to reduce redundancy, enforce relationships through foreign keys, and allow for flexible querying in Oracle APEX.

**Database Implementation**

To implement the database into Oracle APEX, we wrote four CREATE TABLE commands, one for each table in the relational schema. The parent tables were INDUSTRY and SKILLS, so we created those first to ensure that the child tables could properly reference them:

INDUSTRY:

CREATE TABLE INDUSTRY (

IndustryID NUMBER,

IndustryName VARCHAR2(100),

CONSTRAINT pk\_industry PRIMARY KEY (IndustryID)

);

SKILLS:

CREATE TABLE SKILLS (

SkillID NUMBER,

SkillName VARCHAR2(100),

CONSTRAINT pk\_skills PRIMARY KEY (SkillID)

);

We then created the two child tables, JOBS and JOB\_SKILL, which reference the parent tables via foreign keys:

JOBS:

CREATE TABLE JOBS (

JobID NUMBER,

JobTitle VARCHAR2(100),

Company VARCHAR2(100),

Location VARCHAR2(100),

ExperienceLevel VARCHAR2(50),

Salary NUMBER,

IndustryID NUMBER,

CONSTRAINT pk\_jobs PRIMARY KEY (JobID),

CONSTRAINT fk\_jobs\_industry FOREIGN KEY (IndustryID) REFERENCES INDUSTRY(IndustryID)

);

JOB\_SKILL:

CREATE TABLE JOB\_SKILL (

JobID NUMBER,

SkillID NUMBER,

CONSTRAINT pk\_job\_skill PRIMARY KEY (JobID, SkillID),

CONSTRAINT fk\_jobskill\_job FOREIGN KEY (JobID) REFERENCES JOBS(JobID),

CONSTRAINT fk\_jobskill\_skill FOREIGN KEY (SkillID) REFERENCES SKILLS(SkillID)

);

For each of the tables, we selected appropriate data types that matched the original dataset. We used the APEX Data Upload tool to upload the clean and normalized data. Below, we have example INSERT commands for each table:

INSERT INTO INDUSTRY (IndustryID, IndustryName)

VALUES (1, ‘Information Technology’);

INSERT INTO SKILLS (SkillID, SkillName)

VALUES (1, ‘Python’)

INSERT INTO JOBS (JobID, JobTitle, Company, Location, ExperienceLevel, Salary, IndustryID) VALUES (1, ‘Data Analyst’, ‘TechCorp’, ‘New York’, ‘Mid Level’, 70000, 1);

INSERT INTO JOB\_SKILL (JobID, SkillID)

VALUES (1, 1);

**Analysis**

We created questions that would help showcase how the data could be used and solutions that it could create for a user.

**Question 1: In-Demand Skills**

What are the most in-demand skills across different job postings? What skills are associated with earning the highest salaries? To identify the most in-demand skills among the job postings, we aimed to quantify the frequency of each skill appearing across job listings.

To answer these questions, we needed to count the number of job postings associated with each skill. Since skills are stored separately in the SKILLS table and linked to jobs via the JOB\_SKILL table, we made a query that joins these two tables. We grouped by skill name, which allowed us to count the job postings for each skill, and then sort the results in descending order for the most in-demand skills.

SELECT

s.SkillName,

COUNT (js.JobID) AS JobCount

FROM

SKILLS s

JOIN

JOB\_SKILL js ON s.SkillID = js.SkillID

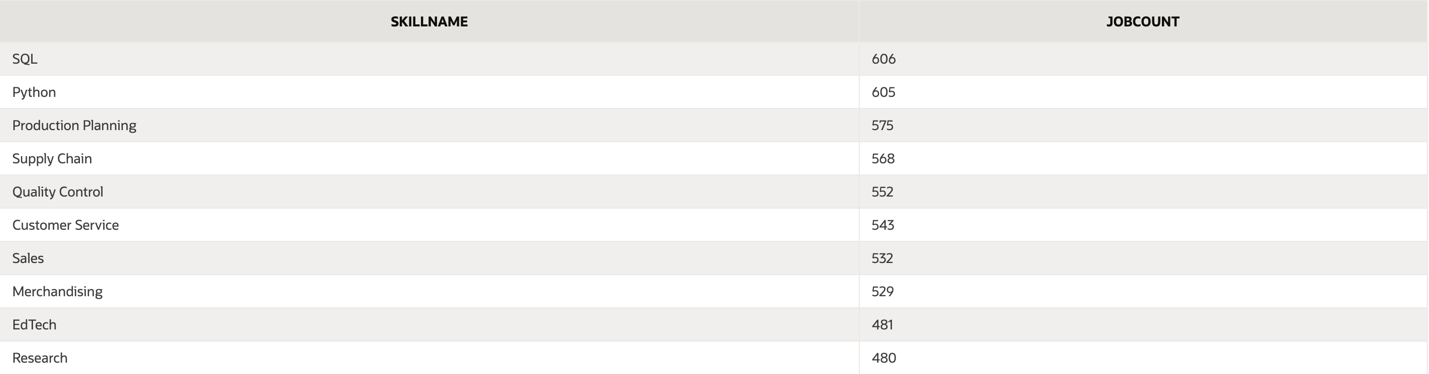
GROUP BY

s.SkillName

ORDER BY

JobCount DESC;

The output of this query revealed that skills related to data analysis and software development are the most frequently required. Specifically, SQL, Python, and Production Planning appeared in the largest number of job listings. These results make a lot of sense, as they align with current job market trends in the workplace. Figure 4 below shows the top results as a table.



*Figure 4 Top 10 In-Demand Skills*

The table above shows the top 10 most in-demand skills as a bar chart, making it easy to compare the frequency of each skill.

That wasn’t enough though, so we made another query to determine which skills are associated with the highest salaries. To do this, we made a query that calculates the average salary for each skill. We combined the SKILLS, JOB\_SKILL, and JOBS tables to link skills to salaries. The query calculates the average salary for each skill while filtering out jobs with missing salary data to ensure accurate results. We then sorted the results by average salary to allow us to see the most lucrative skills.

SELECT

s.SkillName,

COUNT(js.JobID) AS JobCount,

ROUND(AVG(j.Salary), 2) AS AvgSalary

FROM

SKILLS s

JOIN

JOB\_SKILL js ON s.SkillID = js.SkillID

JOIN

JOBS j ON js.JobID = j.JobID

WHERE

j.Salary IS NOT NULL

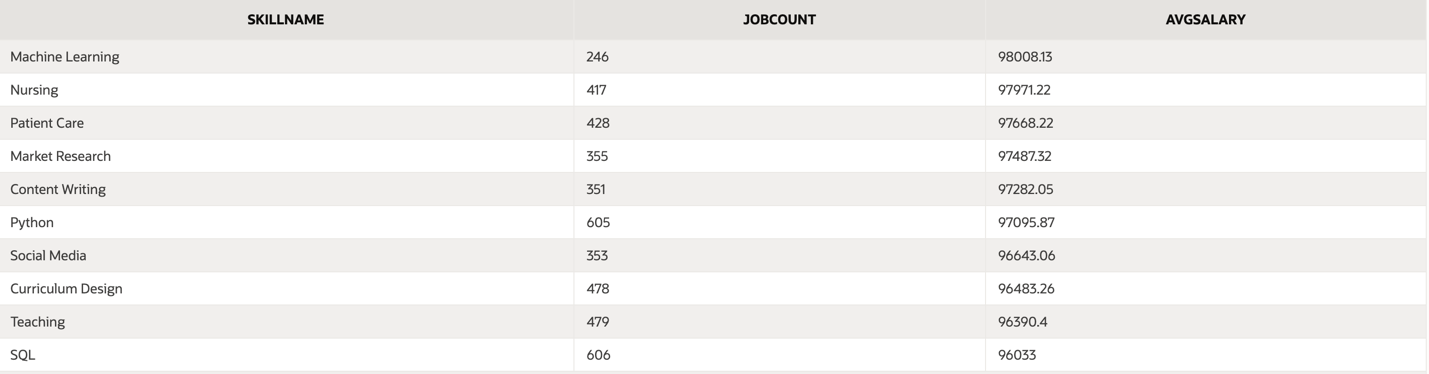
GROUP BY

s.SkillName

ORDER BY

AvgSalary DESC;

The analysis highlighted that skills such as machine learning, nursing, and patient care are linked to the highest average salaries. Behind that, skills such as Python and SQL were in the top 10 results, showing that they are both in demand (popular jobs) and also make a lot of money. The table below (Figure 5) shows a table with the top results from the query.



*Figure 5 Top 10 Average Salaries*

Figure 16 shows the information as a bar chart.

**Question 2: Industry Salaries**

Do certain industries generally offer higher base salaries? This question is especially relevant for students and current workers looking to change careers by gauging the value of entering a certain industry over another. While factors such as job function, experience level, location, and company size all play a role in determining compensation, our goal in this analysis is to explore strictly the impact of the industry on employee salary.

To address this question, we joined the jobs table with the industry reference table using the shared key INDUSTRYID. The jobs table contains individual job listings, that each had a specific industry from a foreign key. The industry table has the label (INDUSTRYNAME) for each industry ID, allowing us to clearly interpret the results.

Before comparing salaries across industries, we ran a query to count how many job listings existed in each industry. This helped ensure that our analysis wouldn’t be skewed by industries with very few records. The query joined the jobs and industry tables and grouped the results by industry name. By verifying that each industry had a sufficient sample size, we increased the reliability and validity of our salary comparisons.

SELECT

i.INDUSTRYNAME,

COUNT(j.JOBID) AS JobCount

FROM

jobs j

JOIN

industry i ON j.INDUSTRYID = i.INDUSTRYID

GROUP BY

i.INDUSTRYNAME

ORDER BY

JobCount DESC;

The results of that query are shown below in figure 6

A table with numbers and text

AI-generated content may be incorrect.

*Figure 6 Industry Job Count*

Once we verified that there were enough job listings from each industry that the tests would be accurate, we then wrote an aggregate query to group the joined data by INDUSTRYNAME and calculate the average base salary (SALARY) within each industry. The output was sorted in descending order to highlight the highest paying industries at the top of the list and the lowest paying ones at the bottom.

SELECT

i.INDUSTRYNAME,

ROUND(AVG(j.SALARY), 2) AS AVERAGE\_SALARY

FROM

jobs j

JOIN

industry i ON j.INDUSTRYID = i.INDUSTRYID

GROUP BY

i.INDUSTRYNAME

ORDER BY

AVERAGE\_SALARY DESC;

The results of this query are shown in figure 7

A table with numbers and text

AI-generated content may be incorrect.

*Figure 7 Average Salary by Industry*

The data does show variation in average base salary across different industries. **Marketing** emerged as the industry with the highest average base salary, followed closely by **Healthcare**, **Finance**, and **Retail**. These industries reported mean salaries above $95,000, suggesting that roles within these sectors tend to offer more competitive compensation packages. At the other end of the spectrum, **Manufacturing** and **Software** ranked lowest in average salary. I found this to be interesting as I initially assumed software would be one of the highest paying industries given current market trends.

These findings have important implications. For students pursuing careers in business, data analytics, or IT, the choice of industry may matter nearly as much as the job function itself. For instance, a business analyst role in the healthcare or marketing sector may yield higher pay than a similar role in education or nonprofit work. While these numbers certainly reflect a clear difference between sectors, the range of averages is still $3,093.45, which for some people may be a significant enough difference to completely change industry fields, but for others the pay difference could be seen as insignificant enough to pick a lower paying industry out of a liking for the field.

Overall, the analysis demonstrates that industry does in fact play a significant role in determining salary, a key detail for anyone entering the job market.

**Question 3: Salaries across varying geographical locations**

How do salaries vary across different geographical locations within the dataset? To address this question, we composed a SQL query that summarizes salary statistics by city. Specifically, the query grouped the data by the LOCATION field and calculated the number of job listings in each city, along with the average, minimum, and maximum salaries offered

SELECT LOCATION, COUNT(\*) AS JOB\_COUNT, TO\_CHAR(ROUND(AVG(SALARY), 2), '$999,999.00')

AS AVERAGE\_SALARY, TO\_CHAR(MIN(SALARY), '$999,999.00') AS MINIMUM\_SALARY,

TO\_CHAR(MAX(SALARY), '$999,999.00') AS MAXIMUM\_SALARY

FROM JOBS,

GROUP BY LOCATION,

ORDER BYAVG(SALARY) DESC;

Figure 8 shows that London reported the highest average salary ($96,894.95), followed closely by Toronto, San Francisco, and Berlin. Cities such as Bangalore and Sydney showed lower average salaries. Interestingly, the minimum and maximum salaries were identical across all cities, suggesting the dataset may be capped by predefined salary limits or influenced by outliers.

*Figure 8, average salary by city*

To deepen our understanding of what drives average salary differences, we ran a follow-up query that breaks down job counts by experience level within each city:

SELECT LOCATION, SUM(CASE WHEN EXPERIENCELEVEL = 'Senior Level' THEN 1 ELSE 0

END) AS SENIOR\_COUNT, SUM(CASE WHEN EXPERIENCELEVEL = 'Mid Level' THEN 1 ELSE

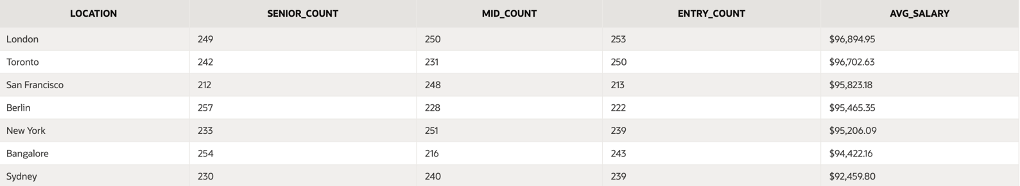
0 END) AS MID\_COUNT,SUM(CASE WHEN EXPERIENCELEVEL = 'Entry Level' THEN 1 ELSE

0 END) AS ENTRY\_COUNT, TO\_CHAR(ROUND(AVG(SALARY), 2),'$999,999.00') AS AVG\_SALARY

FROM JOBS

GROUP BY LOCATION

ORDER BY AVG\_SALARY DESC;

The results (see Figure 9) reveal that cities with a higher proportion of senior-level positions tend to also report higher average salaries

*Figure 9*

From this analysis, we found that experience level is a strong predictor of salary. London, Berlin, and Toronto have a high number of senior-level jobs, which aligns with their top salary rankings. Mid-level roles also play a significant role, especially in cities like San Francisco and New York, where such positions are common and well-compensated due to local market rates. Bangalore and Sydney, while also showing many senior-level roles, report lower average salaries—likely a reflection of regional wage standards or currency differences. The uniform salary range ($40,000 to $150,000) across all locations limits our ability to identify true outliers or high-earning roles, which may indicate normalization of the data or fixed salary boundaries within the dataset.

**Question 4: Industry Growth Over Time**

Which industries are seeing the most growth in job availability? This question is key for students and job seekers who want to set their career paths towards growing job markets. While many factors drive industry trends, this analysis focuses specifically on the increase in job postings within each industry across the dataset’s timeframe to help identify growing fields.

To show this, we wrote a SQL query that split the job postings dataset into two time sections: early and late. We did this by assigning a ‘period’ label based on whether each job entry’s ID fell below or above the median ID (the midpoint of the dataset). By joining job data with the industry table and grouping by industry and period, we were able to calculate the number of postings in each phase. This let us determine the change in job volume across time.

To highlight industries with the most growth, we subtracted early job counts from late job counts for each industry. The results (shown in Figure 10) revealed that **Marketing** experienced the highest growth in job postings, increasing by 31 listings. **Finance** followed with 22 new postings, while other sectors like **Software** and **Manufacturing** showed modest gains. On the other hand, **Education** and **Retail** experienced declines, suggesting reduced hiring momentum in those areas.

These trends provide helpful insights for students exploring career paths. Growth in job postings can indicate rising demand, which may signal better job availability or more opportunities in those fields. While job function and qualifications are still key factors, knowing which industries are actively expanding can help guide more informed career decisions.

WITH job\_data AS (

SELECT

j.ID AS job\_id,

i.INDUSTRYNAME

FROM jobs j

JOIN industry i ON j.INDUSTRYID = i.INDUSTRYID

),

job\_with\_timeframe AS (

SELECT

INDUSTRYNAME,

job\_id,

CASE

WHEN job\_id <= (SELECT MAX(ID) FROM jobs) / 2 THEN 'early'

ELSE 'late'

END AS period

FROM job\_data

),

industry\_growth AS (

SELECT

INDUSTRYNAME,

period,

COUNT(\*) AS job\_count

FROM job\_with\_timeframe

GROUP BY INDUSTRYNAME, period

)

SELECT

ig1.INDUSTRYNAME,

ig1.job\_count AS early\_postings,

ig2.job\_count AS late\_postings,

(ig2.job\_count - ig1.job\_count) AS growth

FROM

industry\_growth ig1

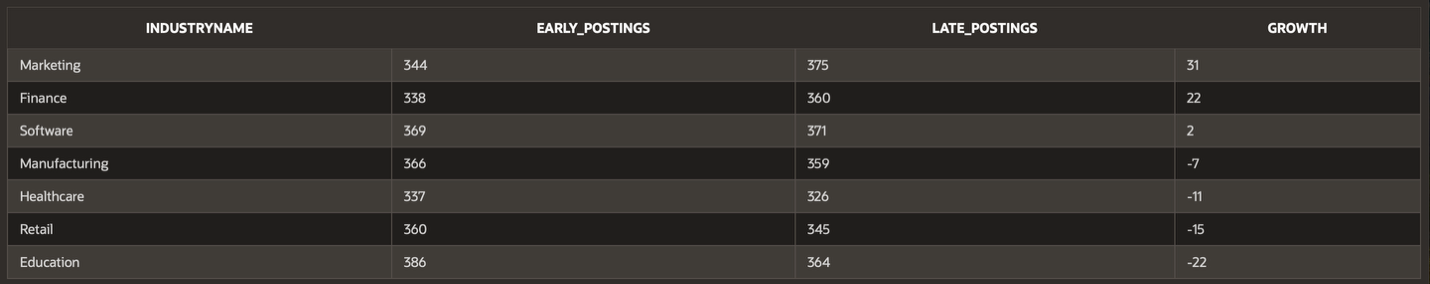
JOIN

industry\_growth ig2 ON ig1.INDUSTRYNAME = ig2.INDUSTRYNAME

WHERE

ig1.period = 'early' AND ig2.period = 'late'

ORDER BY growth DESC;



*Figure 10*

**Question 5: Salary by Experience Level**

What role does experience level play in determining salary for job postings? This question addresses common assumptions that more experience directly correlates with higher pay. To expand on this, we analyzed the average salaries across three experience categories Entry Level, Mid Level, and Senior Level based on the survey data.

SELECT

EXPERIENCELEVEL AS label,

ROUND(AVG(SALARY), 2) AS value

FROM

JOBS

GROUP BY

EXPERIENCELEVEL

ORDER BY

value DESC;

Surprisingly, Entry Level positions reported the highest average salary at $96,012, slightly exceeding Mid Level ($95,370) and Senior Level ($94,512) roles.

This unexpected trend may reflect a dataset skew toward technical or in demand Entry Level roles that command higher wages despite requiring less formal experience. On top of that, the data may not fully capture job complexity or education requirements, which often influence compensation. These results show that pay distribution across experience levels in this dataset is relatively flat. These insights are helpful for early-career professionals assessing the financial implications of entry versus advanced roles.

Our results of how average salary varies amongst experience level is displayed below in figure 11.

A table with numbers and text

AI-generated content may be incorrect.

*Figure 11*

**Web Design**

Homepage

The home page of our web application introduces a brief project summary and a professional image that aligns with the theme. A nested navigation menu with custom icons organizes key report sections. A consistent accent color is used throughout to highlight important content and maintain visual appeal, as shown in Figure 12.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 12 Home page*

Tables

We created an interactive report for each of the four main database table. This enables users to dynamically search, filter, and group the data based on relevant criteria. Each report page includes labeled column headings and customized number formats to enhance clarity and usability. Additionally, every page contains a descriptive text box summarizing the purpose and contents of the table, helping users quickly understand the structure and relevance of the data being presented (Figures 13-16).

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 13 Industries*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 14 Jobs*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 15 Skill ID*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 16 Skills*

Queries

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 17 In-Demand Skills*

The first question was about which skills were most popular under the job postings that were in the database (Figure 17). This included a bar chart where we totaled the number of times a skill came up in the dataset, and then we returned the top 10 skills. This page was a classic dashboard report, where we also returned a table showcasing the top salaries by each skill type, which showed little overall variance between the top jobs in the dataset, but machine learning, nursing, and patient care were the top skills.

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 18 Industry Salary Analysis*

The Second question, as represented in figure 18 was used to get a basic understanding of how a certain industry can affect the potential pay in an ai driven field. We first looked at the count of jobs in each industry to ensure that there was an even enough distribution to compare the industries amongst each other. After verifying this through the pie chart, we then used the bar chart to display the variance of salary within different industries.

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 19 Salaries across varying geographical locations*

The third research question explores salary variation across major cities and examines the distribution of job levels within each location. These insights were visualized using both horizontal and vertical bar charts to provide a clear comparison of average salaries and job counts by city (Figure 19).

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 20 Industry growth over time*

This analysis explores which industries experienced the most change in job postings over the dataset’s timeframe by comparing early and late periods. The bar chart shows net growth in postings across several industries. Marketing had the highest growth, followed by Finance and Software, indicating increased demand in those sectors. In contrast, industries like Retail and Education saw significant declines, reflecting a shift away from traditional roles toward more digital and business-focused opportunities.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 21 Salary by experience level*

The final question we had, as seen in figure 21, was how does experience level affect the salary offered. In this analysis, we decided to use a bar chart to compare the three levels amongst each other. This would give us a solid understanding of how experience is treated within ai driven fields.

1. <https://www.kaggle.com/datasets/samayashar/ai-powered-job-recommendations> [↑](#footnote-ref-1)