Project Title: Student Placement Analysis

Objective:

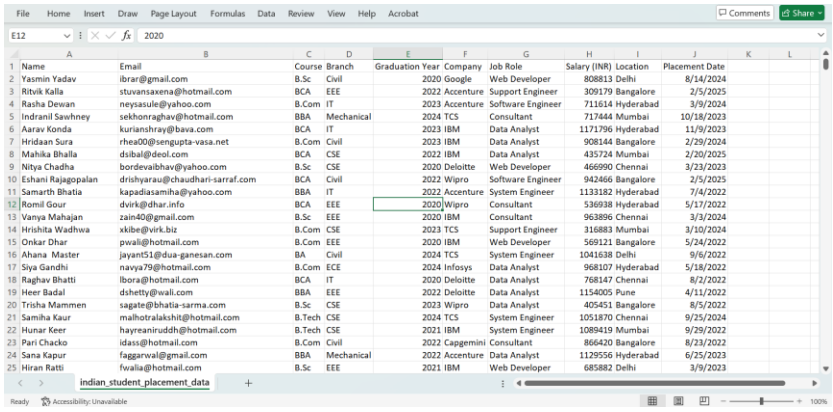
To analyze student placement data to discover hiring patterns, salary distribution, job role trends, and company preferences using data analytics tools. This project provides insights into which branches and companies are leading in campus placements and what salary trends exist across graduation years.

Tools Used:

- **Excel** – Data cleaning, formulas, charting
- **MySQL Workbench** – Database creation, SQL queries
- **Python (Jupyter Notebook)** – Exploratory data analysis
- **Tableau** – Interactive dashboards
- **Libraries:** Pandas, Matplotlib, Seaborn

Dataset Overview:

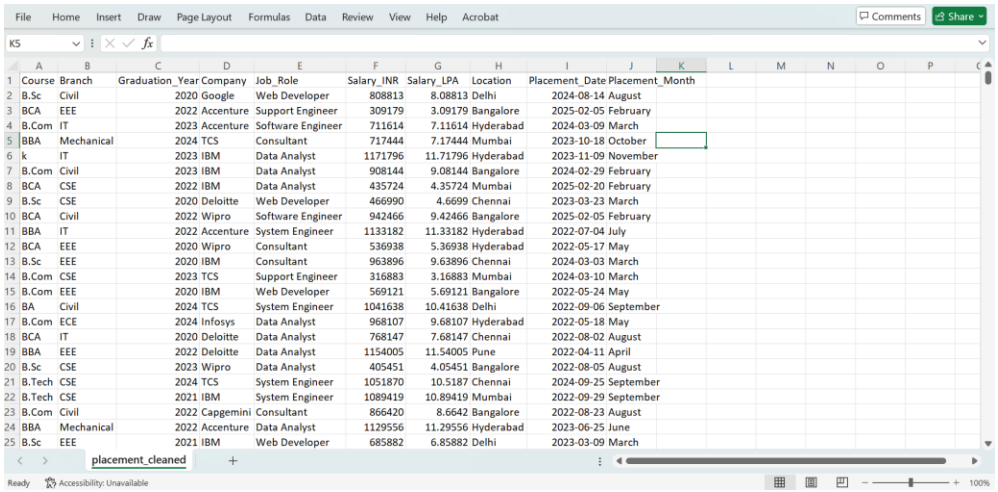
- File: indian_student_placement_data.csv
- Total Records: 1000 students
- Key Fields:
    - Name, Email, Course, Branch, Graduation Year
    - Company, Job Role, Salary (INR)
    - Location, Placement Date, Placement Month



	A	B	C	D	E	F	G	H	I	J
1	Name	Email	Course	Branch	Graduation Year	Company	Job Role	Salary (INR)	Location	Placement Date
2	Yasmin Yadav	ibrar@gmail.com	B.Sc	Civil	2020	Google	Web Developer	808813	Delhi	8/14/2024
3	Ritvik Kalla	stuvansaxena@hotmail.com	BCA	EEE	2022	Accenture	Support Engineer	309179	Bangalore	2/5/2025
4	Rasha Dewan	neysasule@yahoo.com	B.Com	IT	2023	Accenture	Software Engineer	711614	Hyderabad	3/9/2024
5	Indranil Sawhney	sekhonraghav@hotmail.com	BBA	Mechanical	2024	TCS	Consultant	717444	Mumbai	10/18/2023
6	Aarav Konda	kurianshray@bava.com	BCA	IT	2023	IBM	Data Analyst	1171796	Hyderabad	11/9/2023
7	Hridaan Sura	rhea00@sengupta-vasa.net	B.Com	Civil	2023	IBM	Data Analyst	908144	Bangalore	2/29/2024
8	Mahika Bhalla	dsibal@deol.com	BCA	CSE	2022	IBM	Data Analyst	435724	Mumbai	2/20/2025
9	Nitya Chadha	bordevaibhav@yahoo.com	B.Sc	CSE	2020	Deloitte	Web Developer	466990	Chennai	3/23/2023
10	Eshani Rajagopalan	drishyarau@chaudhari-sarraf.com	BCA	Civil	2022	Wipro	Software Engineer	942466	Bangalore	2/5/2025
11	Samarth Bhatia	kapadiasamiha@yahoo.com	BBA	IT	2022	Accenture	System Engineer	1133182	Hyderabad	7/4/2022
12	Romil Gour	dvirk@dhar.info	BCA	EEE	2020	Wipro	Consultant	536938	Hyderabad	5/17/2022
13	Vanya Mahajan	zain40@gmail.com	B.Sc	EEE	2020	IBM	Consultant	963896	Chennai	3/3/2024
14	Hrishita Wadhwa	xkibe@virk.biz	B.Com	CSE	2023	TCS	Support Engineer	316883	Mumbai	3/10/2024
15	Onkar Dhar	pwali@hotmail.com	B.Com	EEE	2020	IBM	Web Developer	569121	Bangalore	5/24/2022
16	Ahana Master	jayant51@dua-ganesan.com	BA	Civil	2024	TCS	System Engineer	1041638	Delhi	9/6/2022
17	Siya Gandhi	navya79@hotmail.com	B.Com	ECE	2024	Infosys	Data Analyst	968107	Hyderabad	5/18/2022
18	Raghav Bhatti	lbora@hotmail.com	BCA	IT	2020	Deloitte	Data Analyst	768147	Chennai	8/2/2022
19	Heer Badal	dshetty@wali.com	BBA	EEE	2022	Deloitte	Data Analyst	1154005	Pune	4/11/2022
20	Trisha Mammen	sagate@bhatia-sarma.com	B.Sc	CSE	2023	Wipro	Data Analyst	405451	Bangalore	8/5/2022
21	Samiha Kaur	malhotralakshit@hotmail.com	B.Tech	CSE	2024	TCS	System Engineer	1051870	Chennai	9/25/2024
22	Hunar Keer	hayreaniruddh@hotmail.com	B.Tech	CSE	2021	IBM	System Engineer	1089419	Mumbai	9/29/2022
23	Pari Chacko	idass@hotmail.com	B.Com	Civil	2022	Capgemini	Consultant	866420	Bangalore	8/23/2022
24	Sana Kapur	faggarwal@gmail.com	BBA	Mechanical	2022	Accenture	Data Analyst	1129556	Hyderabad	6/25/2023
25	Hiran Ratti	fwalia@hotmail.com	B.Sc	EEE	2021	IBM	Web Developer	685882	Delhi	3/9/2023

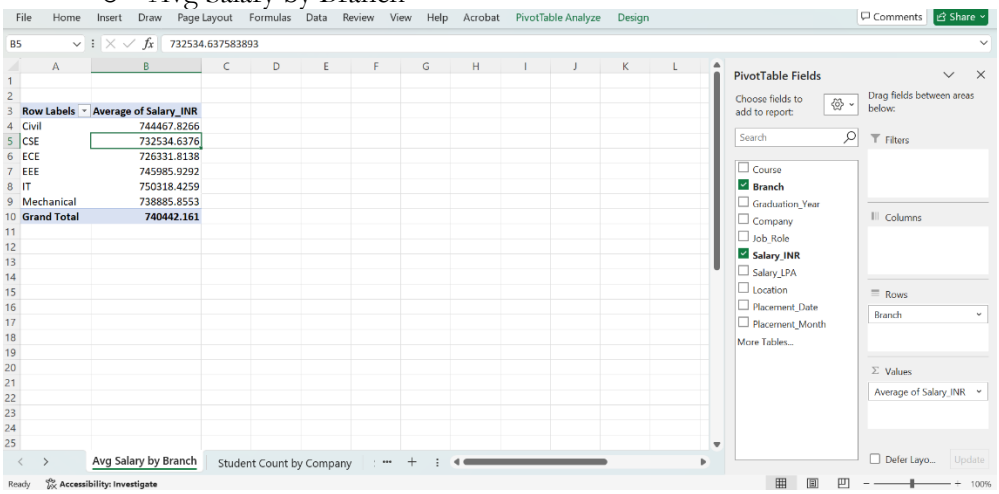
Excel Work Summary:

- File: cleaned_file_with_formulas.xlsx
- Actions performed:
    - Removed unwanted columns (Name, Email)
    - Created derived columns:
        - Salary_LPA = Salary_INR/100000
        - Placement Month = TEXT(Placement_Date, "mmmm")
    - Used conditional formatting on top 10 salaries



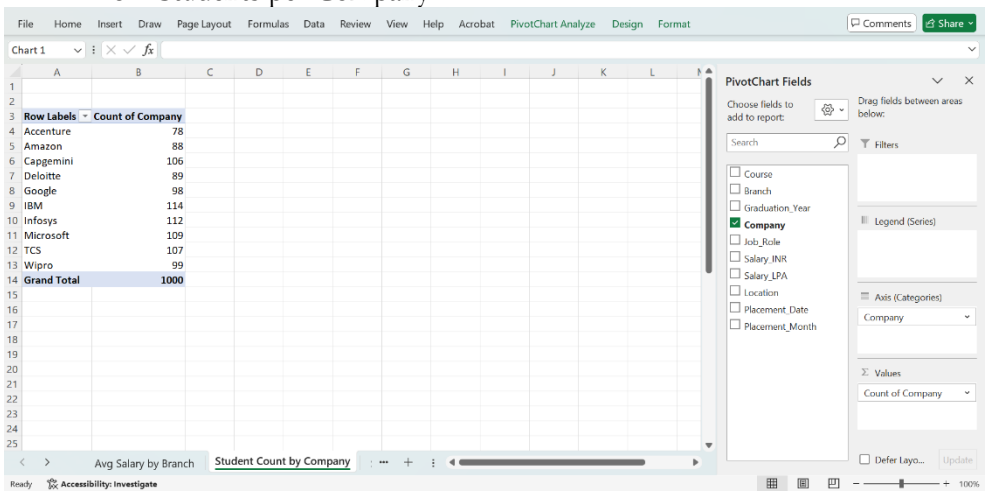
Course	Branch	Graduation_Year	Company	Job_Role	Salary_INR	Salary_LPA	Location	Placement_Date	Placement_Month
B.Sc	Civil	2020	Google	Web Developer	808813	8.08813	Delhi	2024-08-14	August
BCA	EEE	2022	Accenture	Support Engineer	309179	3.09179	Bangalore	2025-02-05	February
B.Com	IT	2023	Accenture	Software Engineer	711614	7.11614	Hyderabad	2024-03-09	March
BBA	Mechanical	2024	TCS	Consultant	717444	7.17444	Mumbai	2023-10-18	October
k	IT	2023	IBM	Data Analyst	1171796	11.71796	Hyderabad	2023-11-09	November
B.Com	Civil	2023	IBM	Data Analyst	908144	9.08144	Bangalore	2024-02-29	February
BCA	CSE	2022	IBM	Data Analyst	435724	4.35724	Mumbai	2025-02-20	February
B.Sc	CSE	2020	Deloitte	Web Developer	466990	4.6699	Chennai	2023-03-23	March
BCA	Civil	2022	Wipro	Software Engineer	942466	9.42466	Bangalore	2025-02-05	February
BBA	IT	2022	Accenture	System Engineer	1133182	11.33182	Hyderabad	2022-07-04	July
BCA	EEE	2020	Wipro	Consultant	536938	5.36938	Hyderabad	2022-05-17	May
B.Sc	EEE	2020	IBM	Consultant	963896	9.63896	Chennai	2024-03-03	March
B.Com	CSE	2023	TCS	Support Engineer	316883	3.16883	Mumbai	2024-03-10	March
B.Com	EEE	2020	IBM	Web Developer	569121	5.69121	Bangalore	2022-05-24	May
BA	Civil	2024	TCS	System Engineer	1041638	10.41638	Delhi	2022-09-06	September
B.Com	ECE	2024	Infosys	Data Analyst	968107	9.68107	Hyderabad	2022-05-18	May
BCA	IT	2020	Deloitte	Data Analyst	768147	7.68147	Chennai	2022-08-02	August
BBA	EEE	2022	Deloitte	Data Analyst	1154005	11.54005	Pune	2022-04-11	April
B.Sc	CSE	2023	Wipro	Data Analyst	405451	4.05451	Bangalore	2022-08-05	August
B.Tech	CSE	2024	TCS	System Engineer	1051870	10.5187	Chennai	2024-09-25	September
B.Tech	CSE	2021	IBM	System Engineer	1089419	10.89419	Mumbai	2022-09-29	September
B.Com	Civil	2022	Capgemini	Consultant	866420	8.6642	Bangalore	2022-08-23	August
BBA	Mechanical	2022	Accenture	Data Analyst	1129556	11.29556	Hyderabad	2023-06-25	June
B.Sc	EEE	2021	IBM	Web Developer	685882	6.85882	Delhi	2023-03-09	March

- Built chart in Budget_Calculator.xlsx with:
    - Avg Salary by Branch



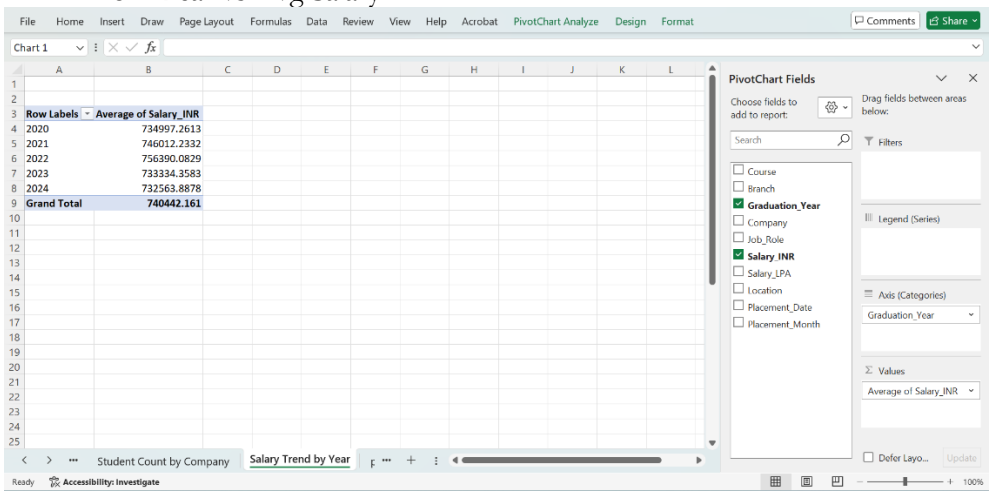
Row Labels	Average of Salary_INR
Civil	744467.8266
CSE	732534.6376
ECE	726331.8138
EEE	745985.9292
IT	750318.4259
Mechanical	738885.8553
Grand Total	740442.161

- Students per Company



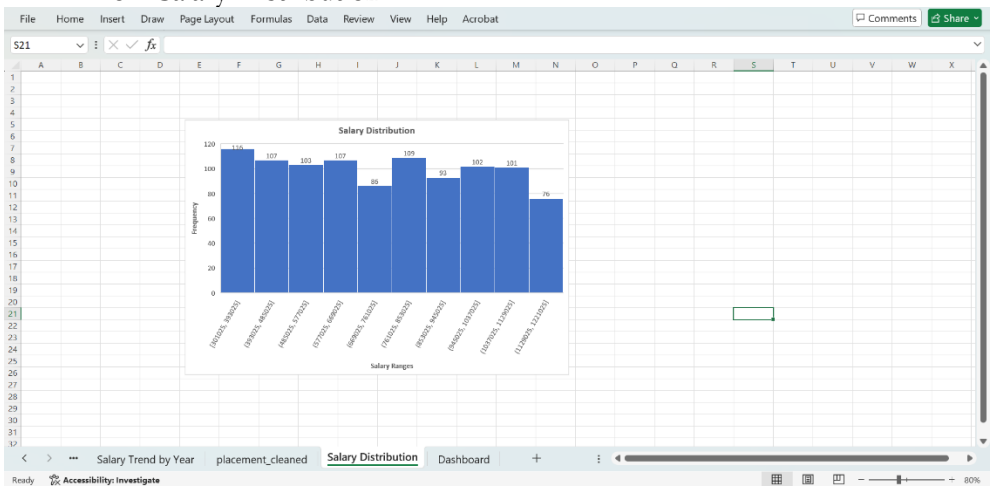
Row Labels	Count of Company
Accenture	78
Amazon	88
Capgemini	106
Deloitte	89
Google	98
IBM	114
Infosys	112
Microsoft	109
TCS	107
Wipro	99
Grand Total	1000

- Year vs Avg Salary

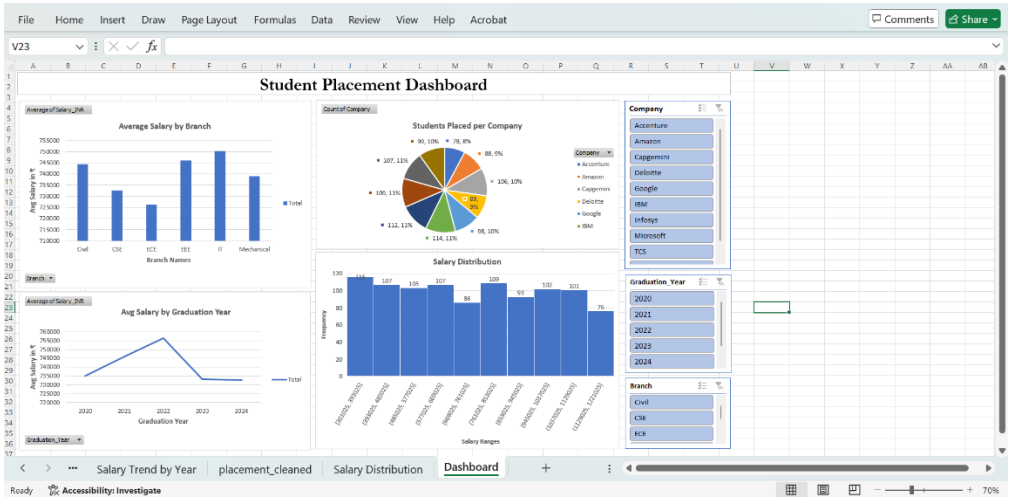


Row Labels	Average of Salary_INR
2020	734997.2613
2021	746012.2332
2022	756390.0829
2023	733334.3583
2024	732563.8878
Grand Total	740442.161

- Salary Distribution



- Final Dashboard



SQL Work Summary:

CREATE DATABASE placement_db;

USE placement_db;

- Imported placement_cleaned.csv using MySQL Workbench

Planned to run queries (to be added in placement_queries.sql) like:

-- Branch-wise Average Salary

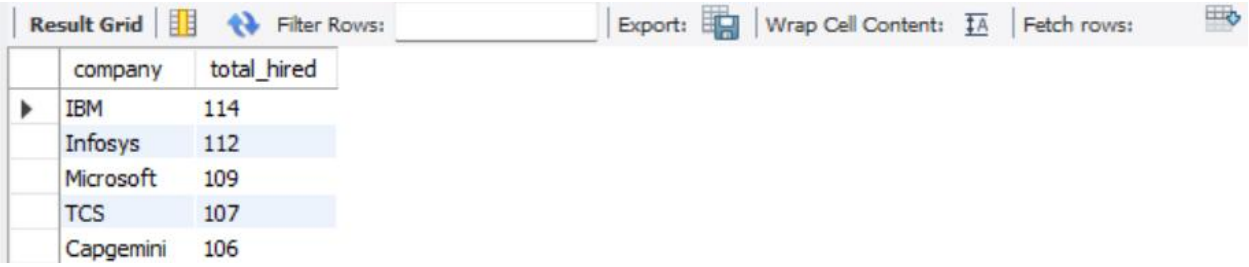
```
SELECT branch, ROUND(AVG(salary_inr)) AS avg_salary
FROM placement_db.placement_cleaned
GROUP BY branch
ORDER BY avg_salary DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	avg_salary
IT	750318
EEE	745986
Civil	744468
Mechanical	738886
CSE	732535
ECE	726332

-- Top 5 Hiring Companies

```
SELECT company, COUNT(*) AS total_hired
FROM placement_db.placement_cleaned
GROUP BY company
ORDER BY total_hired DESC
LIMIT 5;
```

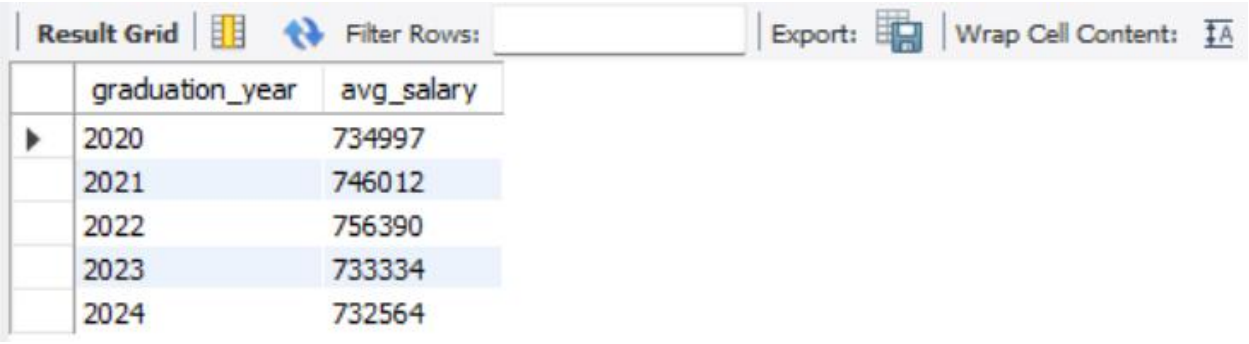


Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

company	total_hired
IBM	114
Infosys	112
Microsoft	109
TCS	107
Capgemini	106

-- Graduation Year vs Avg Salary

```
SELECT graduation_year, ROUND(AVG(salary_inr)) AS avg_salary
FROM placement_db.placement_cleaned
GROUP BY graduation_year
ORDER BY graduation_year;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

graduation_year	avg_salary
2020	734997
2021	746012
2022	756390
2023	733334
2024	732564

-- Most Common Job Roles

```
SELECT job_role, COUNT(*) AS total_placed
FROM placement_db.placement_cleaned
GROUP BY job_role
ORDER BY total_placed DESC
LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

job_role	total_placed
Support Engineer	176
System Engineer	172
Data Analyst	169
Web Developer	165
Consultant	160
Software Engineer	158

-- Students Placed per Location

```
SELECT location, COUNT(*) AS total_placed
FROM placement_db.placement_cleaned
GROUP BY location
ORDER BY total_placed DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

location	total_placed
Bangalore	190
Delhi	175
Chennai	168
Hyderabad	162
Pune	161
Mumbai	144

-- Highest Salary Offered

```
SELECT *
FROM placement_db.placement_cleaned
ORDER BY salary_inr DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Course	Branch	Graduation_Year	Company	Job_Role	Salary_INR	Salary_LPA	Location	Placement_Date	Placement
B.Com	Mechanical	2021	Amazon	Software Engineer	1198129	11.98129	Bangalore	2023-10-20	October

-- Lowest Salary Offered

```
SELECT *
FROM placement_db.placement_cleaned
ORDER BY salary_inr ASC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Course	Branch	Graduation_Year	Company	Job_Role	Salary_INR	Salary_LPA	Location	Placement_Date	Placement_Mont
BA	ECE	2023	Capgemini	System Engineer	301025	3.01025	Delhi	2022-04-27	April

-- Company-wise Average Salary

```
SELECT company, ROUND(AVG(salary_inr)) AS avg_salary
FROM placement_db.placement_cleaned
GROUP BY company
ORDER BY avg_salary DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

company	avg_salary
Capgemini	783987
Accenture	773480
Microsoft	749229
Deloitte	744702
Amazon	739402
Infosys	733734

-- Branch-wise Total Placements

```
SELECT branch, COUNT(*) AS total_placed
FROM placement_db.placement_cleaned
GROUP BY branch
ORDER BY total_placed DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

branch	total_placed
EEE	212
Civil	173
IT	162
Mechanical	159
CSE	149
ECE	145

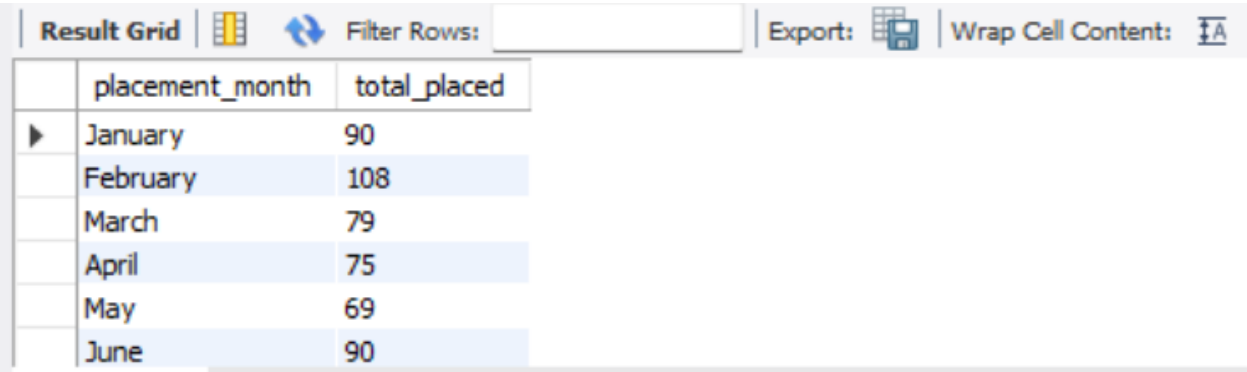
-- Monthly Placement Trend

SELECT placement_month, COUNT(*) AS total_placed

FROM placement_db.placement_cleaned

GROUP BY placement_month

ORDER BY FIELD(placement_month,
'January','February','March','April','May','June','July','August','September','October','November','December');



placement_month	total_placed
January	90
February	108
March	79
April	75
May	69
June	90

Python Work Summary:

- Notebook: placement_eda.ipynb

import data

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

sns.set(style="whitegrid")

load data

df = pd.read_csv(r'E:/student_placement_project/data/placement_cleaned.csv')

df.head()

	Course	Branch	Graduation_Year	Company	Job_Role	Salary_INR	Salary_LPA	Location	Placement_Date	Placement_Month
0	B.Sc	Civil	2020	Google	Web Developer	808813	8.08813	Delhi	2024-08-14	August
1	BCA	EEE	2022	Accenture	Support Engineer	309179	3.09179	Bangalore	2025-02-05	February
2	B.Com	IT	2023	Accenture	Software Engineer	711614	7.11614	Hyderabad	2024-03-09	March
3	BBA	Mechanical	2024	TCS	Consultant	717444	7.17444	Mumbai	2023-10-18	October
4	k	IT	2023	IBM	Data Analyst	1171796	11.71796	Hyderabad	2023-11-09	November

basic overview

Basic Info

df.info()

Summary Statistics

df.describe()

Check for null values

df.isnull().sum()

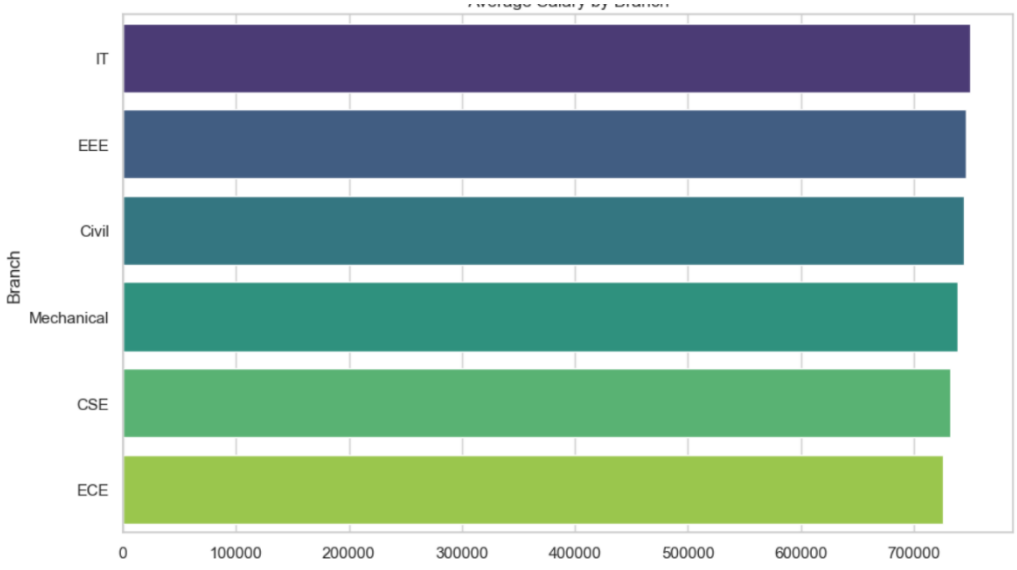
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Course           1000 non-null   object 
 1   Branch           1000 non-null   object 
 2   Graduation_Year  1000 non-null   int64  
 3   Company          1000 non-null   object 
 4   Job_Role         1000 non-null   object 
 5   Salary_INR       1000 non-null   int64  
 6   Salary_LPA       1000 non-null   float64
 7   Location         1000 non-null   object 
 8   Placement_Date   1000 non-null   object 
 9   Placement_Month  1000 non-null   object 
dtypes: float64(1), int64(2), object(7)
memory usage: 78.3+ KB
```

```
Course             0
Branch             0
Graduation_Year    0
Company            0
Job_Role           0
Salary_INR         0
Salary_LPA         0
Location           0
Placement_Date     0
Placement_Month    0
dtype: int64
```

Conducted exploratory data analysis (EDA) to uncover hidden patterns

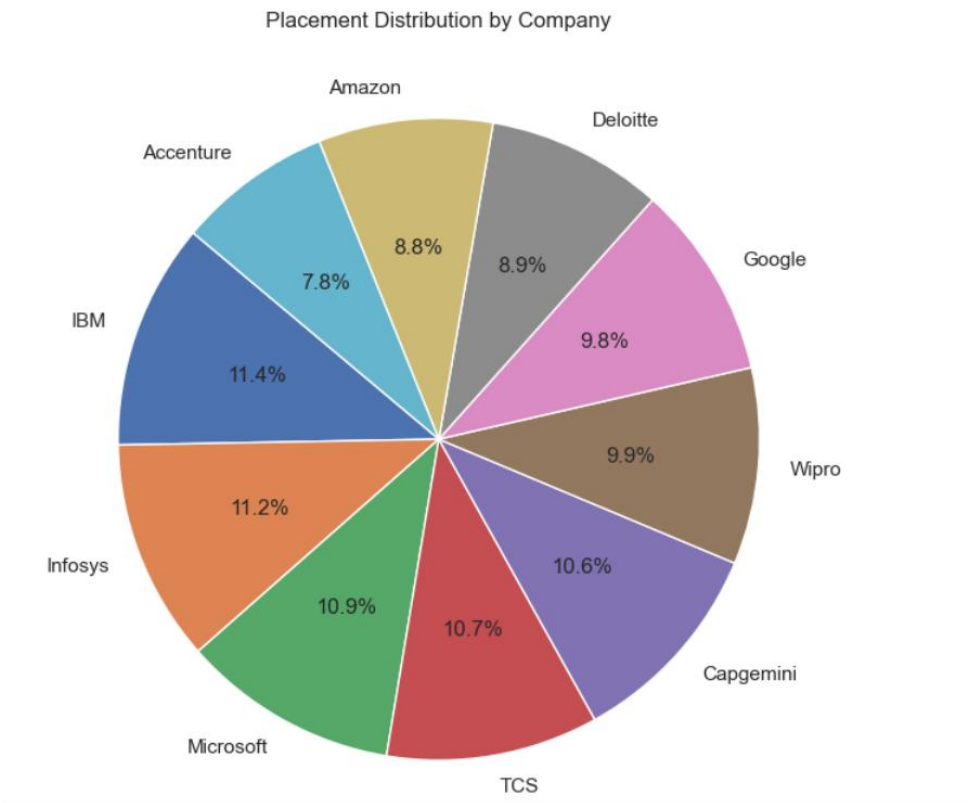
```
# Branch-wise Average Salary
branch_salary = df.groupby('Branch')['Salary_INR'].mean().sort_values(ascending=False)


plt.figure(figsize=(10,6))
sns.barplot(x=branch_salary.values, y=branch_salary.index, palette='viridis')
plt.title('Average Salary by Branch')
plt.xlabel('Average Salary_INR')
plt.ylabel('Branch')
plt.tight_layout()
plt.show()
```



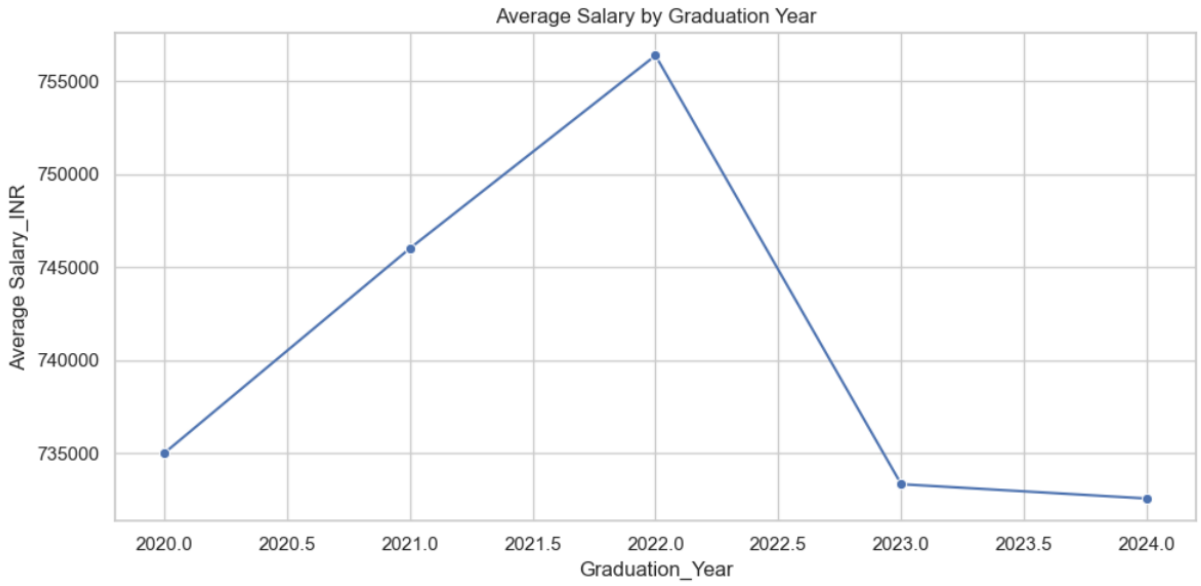
```
# Company-wise Student Count
company_count = df['Company'].value_counts()

plt.figure(figsize=(8,8))
company_count.plot.pie(autopct='%1.1f%%', startangle=140)
plt.title('Placement Distribution by Company')
plt.ylabel('')
plt.show()
```

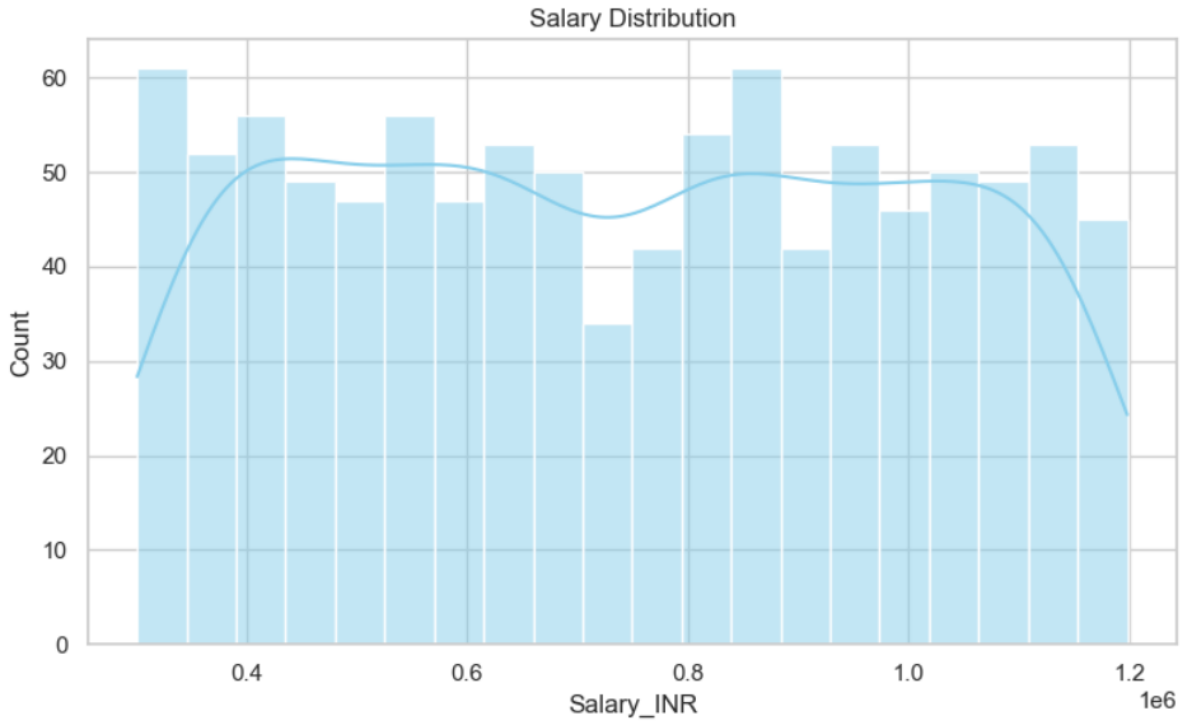


```
# Graduation Year vs Avg Salary
year_salary = df.groupby('Graduation_Year')['Salary_INR'].mean()


plt.figure(figsize=(10,5))
sns.lineplot(x=year_salary.index, y=year_salary.values, marker='o')
plt.title('Average Salary by Graduation Year')
plt.xlabel('Graduation_Year')
plt.ylabel('Average Salary_INR')
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
# Salary Distribution
plt.figure(figsize=(8,5))
sns.histplot(df['Salary_INR'], bins=20, kde=True, color='skyblue')
plt.title('Salary Distribution')
plt.xlabel('Salary_INR')
plt.ylabel('Count')
plt.tight_layout()
plt.show()
```



```
# Placement Month Trend
month_count = df['Placement_Month'].value_counts().reindex([
    'January','February','March','April','May','June','July',
    'August','September','October','November','December'
])

plt.figure(figsize=(10,6))
sns.barplot(x=month_count.index, y=month_count.values, palette='magma')
plt.title('Placements by Month')
plt.xlabel('Month')
plt.ylabel('No. of Placements')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

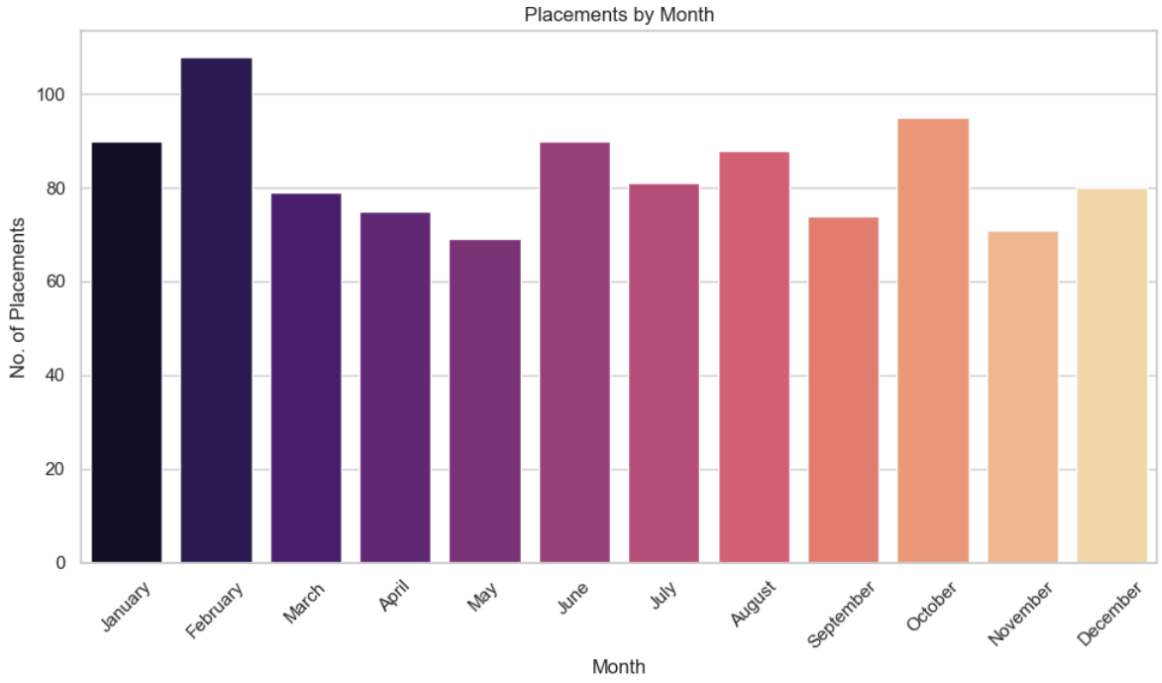
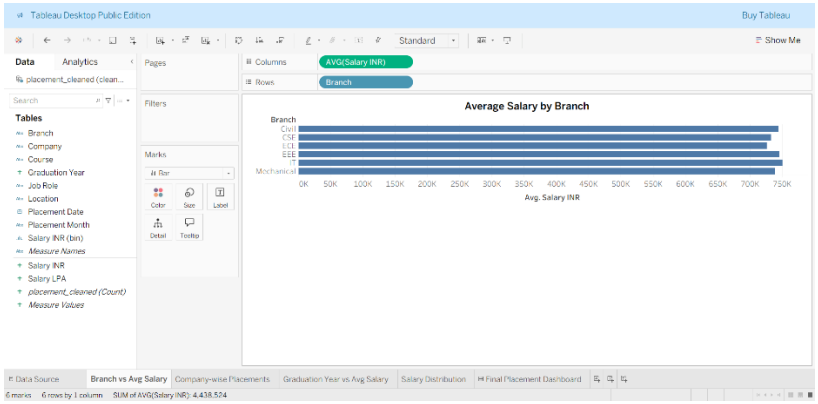
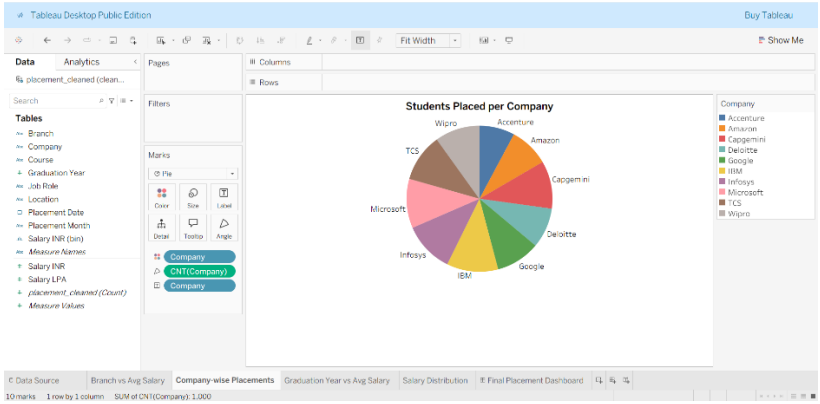


Tableau Work Summary:

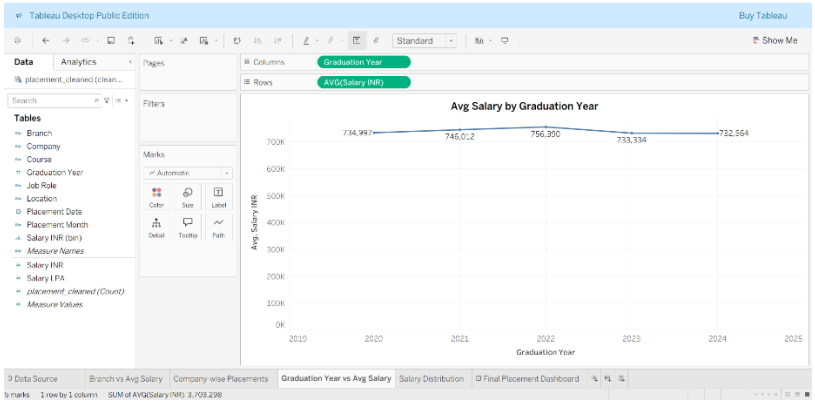
- File: final_dashboard.twbx
- Dashboard includes:
    - **Avg Salary by Branch**



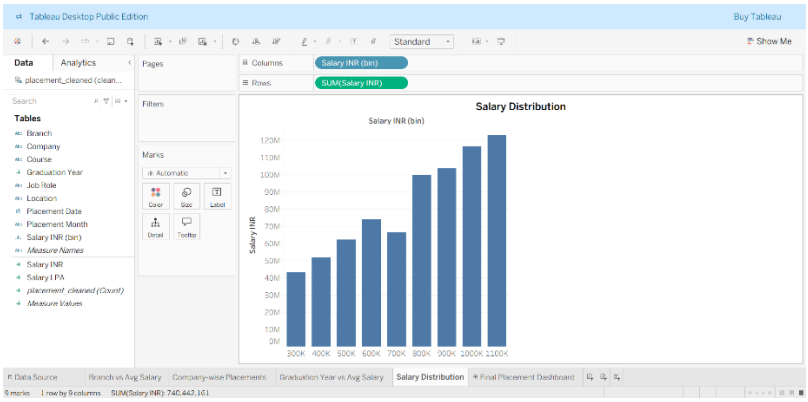
    - **Students Placed per Company**



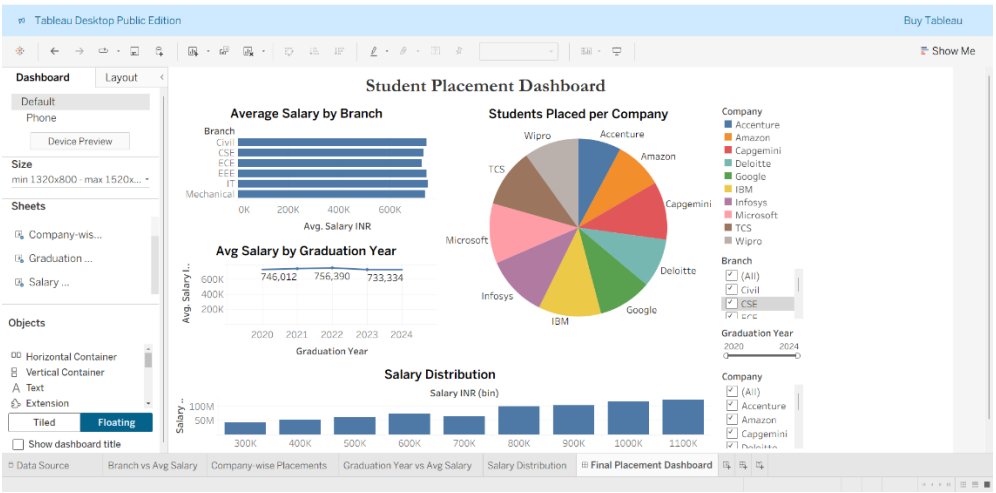
    - **Year vs Salary**



- **Salary Distribution (Histogram)**



- **Final Dashboard**



- Added interactive filters:
  - Branch
  - Graduation Year
  - Company

Insights & Summary:

- **IT** has the highest average salary among all branches
- **IBM** hired the most students overall
- **Graduation Year 2022** had the highest avg salary
- Placements spike in **February and October**
- Majority of salaries fall in the **₹3L to ₹7L** range

Conclusion:

This project demonstrates the full data analysis lifecycle — from raw data cleaning in Excel, structured querying in SQL, EDA in Python, to interactive dashboard building in Tableau.
It uncovers key insights into student placements that can help colleges, training centers, and job seekers understand current hiring patterns and salary expectations. A solid showcase of analytical thinking and tool integration.