Vietnamese - German University

Marking

Project: .....

Python for Data Science

Project title: Graph Plotting in Python

Group members:	Engaging factor (0.0 - 1.0):	Signature:
1. Phan Thị Thúy An	1.0	
2. Hoàng Nhuệ Giang	1.0	
3. Nguyễn Cao Minh Vy	1.0	
4. Dương Trung Hải	1.0	
5. Trần Thùy Trang	1.0	

Group Number: 2

Major: BBA & BFA

Date of submission: 12/05/2023

By signing this page, we declare that all the work in this project is done entirely by our group with no cheating and plagiarism.

Detailed marking scheme

Project (45%)		
Length (5%)	Content (30%)	English (10%)

Comments

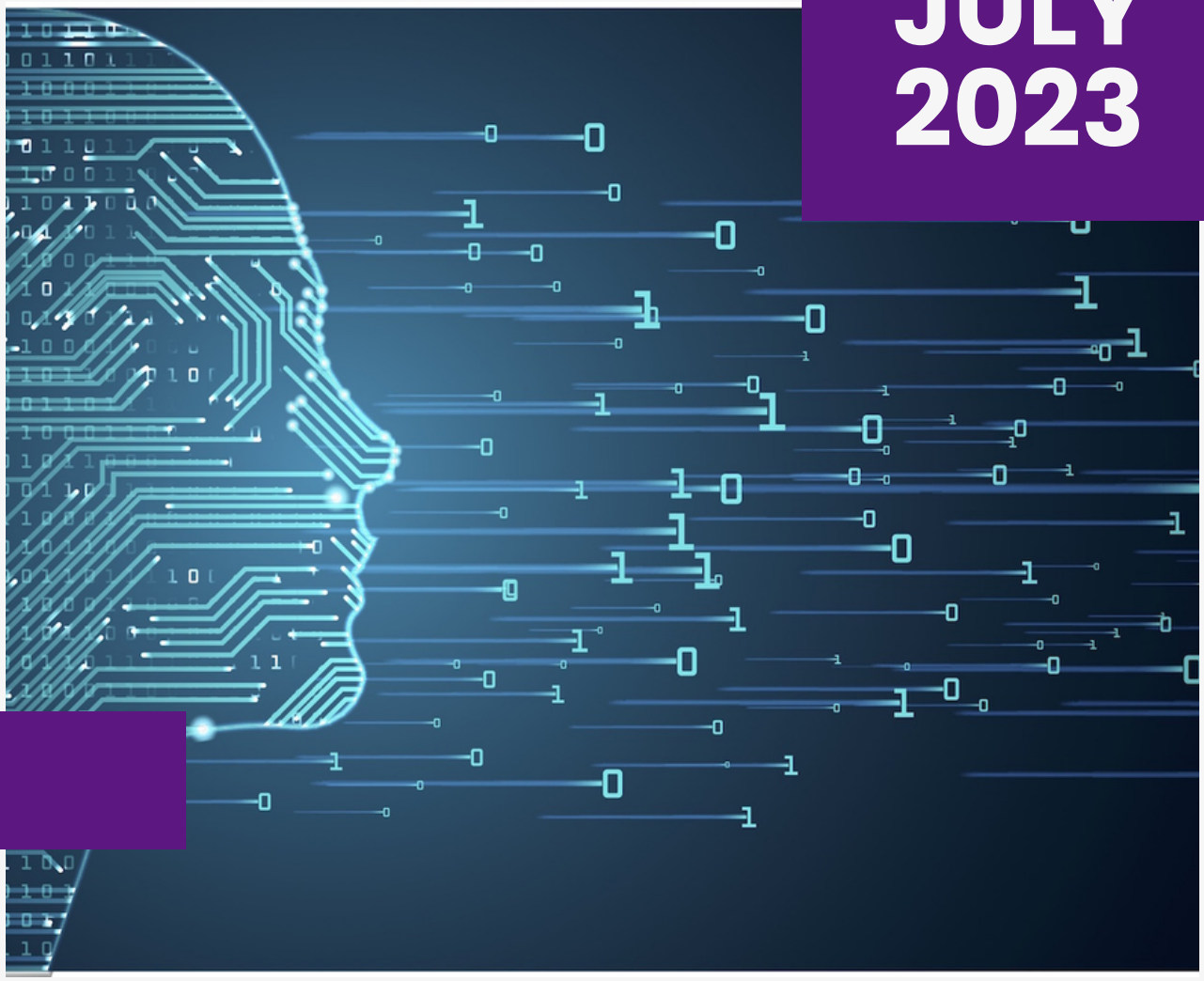
Contribution list

Name	Contribution
Phan Thị Thúy An	Group leader Plot 10, 12, 13
Hoàng Nhuệ Giang	Plot 5, 11, 14
Nguyễn Cao Minh Vy	Plot 2, 7 Editing
Dương Trung Hải	Plot 1, 4, 6, 8
Trần Thùy Trang	Plot 3, 9 Checking grammar

Vietnamese-German University

Vietnamese-German University

JULY
2023



Python Project 1

TUESDAY MORNING CLASS

GROUP 2

Vietnamese-German University

TABLE OF CONTENT

INTRODUCTION

The following report will provide some basic information about the data set, including its author, link, and topic. In addition, it gives 14 plots of this database using the Python program, with an explanation for each plot.

The report was the project for the Python Project 1 of the Business IT 2 Course. The report was prepared by Group 2 in the Tuesday morning class under the guidance and support of Dr. Tan Duc Do.

Thank you very much for taking the time to read the report.

01

02

DATA SET

Campus Recruitment

Academic and Employability Factors influencing placement

- **Author**
  Dr. Dhimant Ganatara
  Senior Manager - Indian Institute of Management, Bangalore
- **Link**
  Campus Recruitment | Kaggle
  https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement
- **What is in it?**
  15 columns & 216 rows
  The data set includes placement data for students at the XYZ campus. It covers secondary and higher secondary school degrees, employability test percentages, and specializations. It also provides the status of placement, work experience, and salary offered to the placed students.

02 DATA SET

- **Why did we choose this data set?**
  Since all members of our group are interested in the process of finding people to work for a company or become new members of an organization, we want to know more about this area. In the process of finding a data set for the R Studio project, we were impressed firstly by the title of this data set, "Campus Recruitment", and secondly by the structure and content of this data file. This data set not only is in the field we enjoy but also has the standard structure we need for this project. And when we study Business IT 2, we are asked to transfer old plots from R Studio to Python.
- **What makes this data collection interesting?**
  This data set includes many factors affecting candidates in recruitment. Also, it has many variables, such as categorical variables and real variables. And we do believe that this diversity can lead to useful data for analysis.
- **What do we want to explore from it?**
  We want to find out the impacts of these factors on each other, such as the relationship between salary and working experience.
  Moreover, we want to find answers to these questions:
  1. How do these given elements matter to a candidate's placement?
  2. Does percentage affect one to get placed?
  3. What degree specialization is in high demand among businesses?

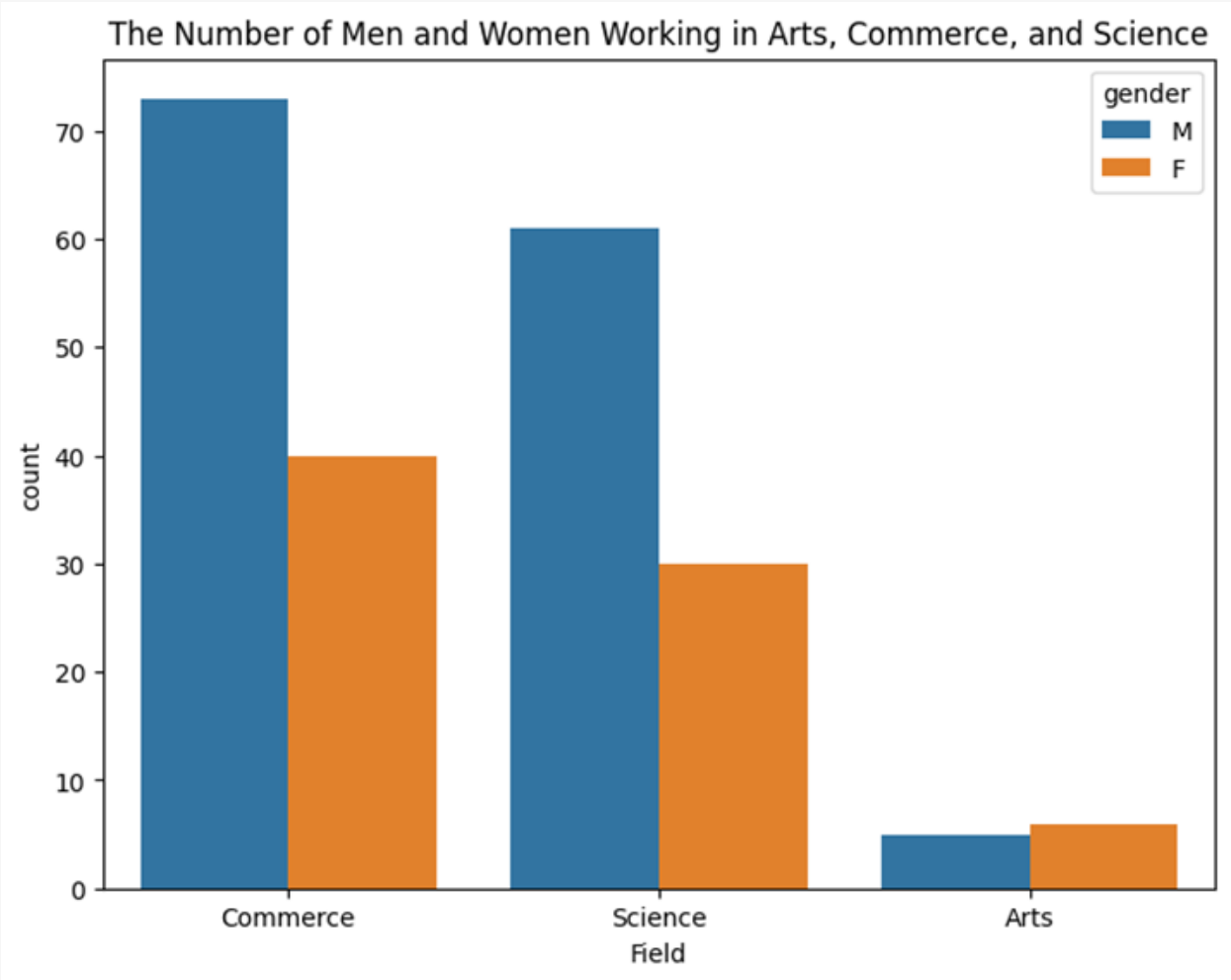
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 1

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
a= pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(8, 6))
sns.countplot(data=a, x='hsc_s', hue='gender')
plt.title("The Number of Men and Women Working in Arts, Commerce, and Science")
plt.xlabel("Field")
plt.show()
```



PLOTS & EXPLANATION

03

Plot 1

The bar graph enumerates the comparison in some certain specializations, such as arts, commerce, and science, between male students and female students in a class at XYZ Campus.

This given chart provides an overview of the gap number of men in the mentioned majors is bigger than that one of women, except for arts, where the difference is negligible.

Based on the diagram, it is apparent that the percentage of women involved in the arts exceeds 5%, whereas the number of men involved is less than 5%. In terms of commerce, over 70% of students identifying as male are engaged in this field, whereas the percentage of female students involved hovers around 40%. Lastly, the field of science sees close to 30% of female participation, while male involvement in this field is nearly double at 60%.

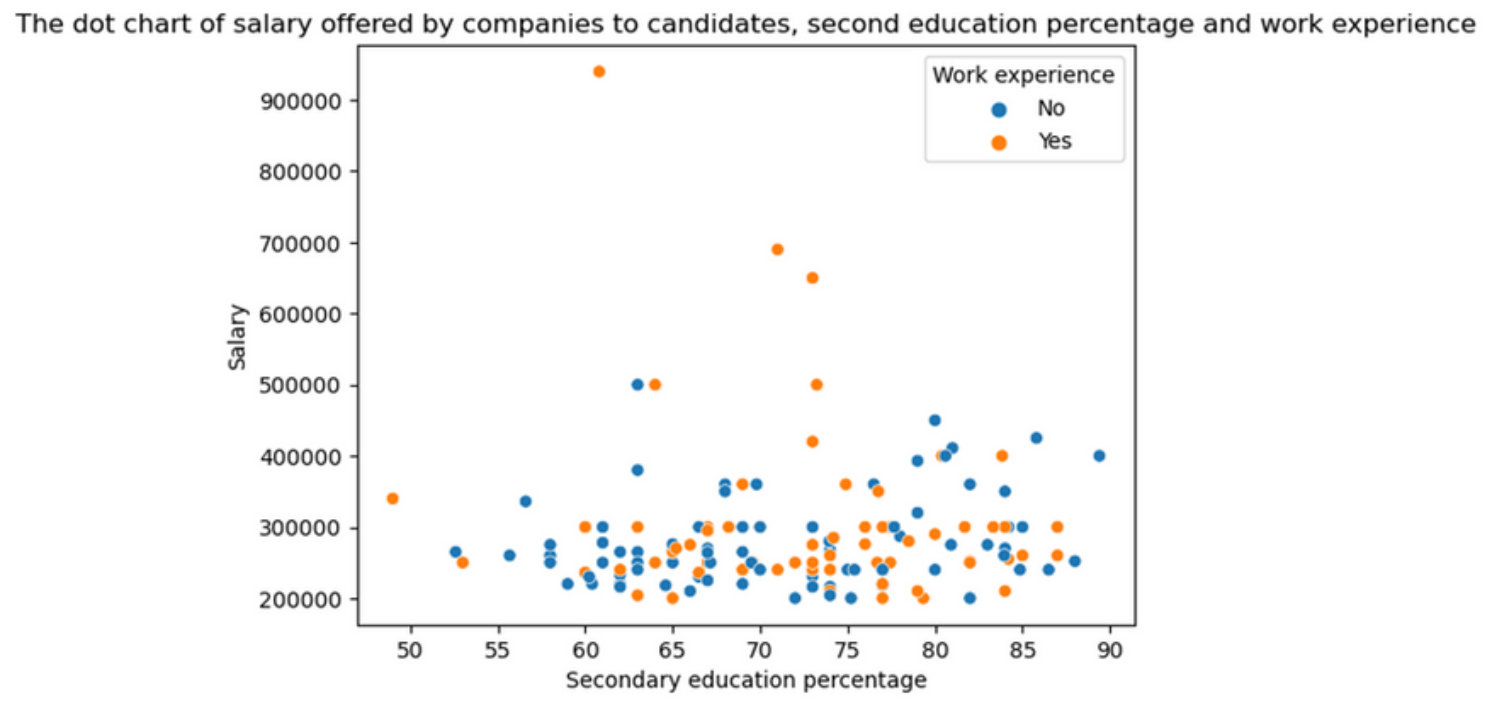
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 2

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
a= pd.read_csv('Placement_Data_Full_Class.csv')
sns.scatterplot(data=a, y='salary',x='ssc_p',hue='workex')
plt.title("The dot chart of salary offered by companies to candidates, second education percentage and work experience")
plt.xlabel("Secondary education percentage")
plt.ylabel("Salary")
plt.legend(title='Work experience')
plt.show()
```



PLOTS & EXPLANATION

03

Plot 2

The provided dot chart reveals the connection of salary offered by companies to candidates, second education percentage and work experience in a campus.

As we can see from the chart, those having worked at somewhere before have higher salary offered than those who have not worked before. Therefore, the highest salary is more than 900000 dollar.

Starting with those are newcomers to the job market, we can observe that the salary range is narrow (from 200000 to 500000 dollar). Therefore, the lowest salary ever experienced in chart is approximately 200000 dollar.

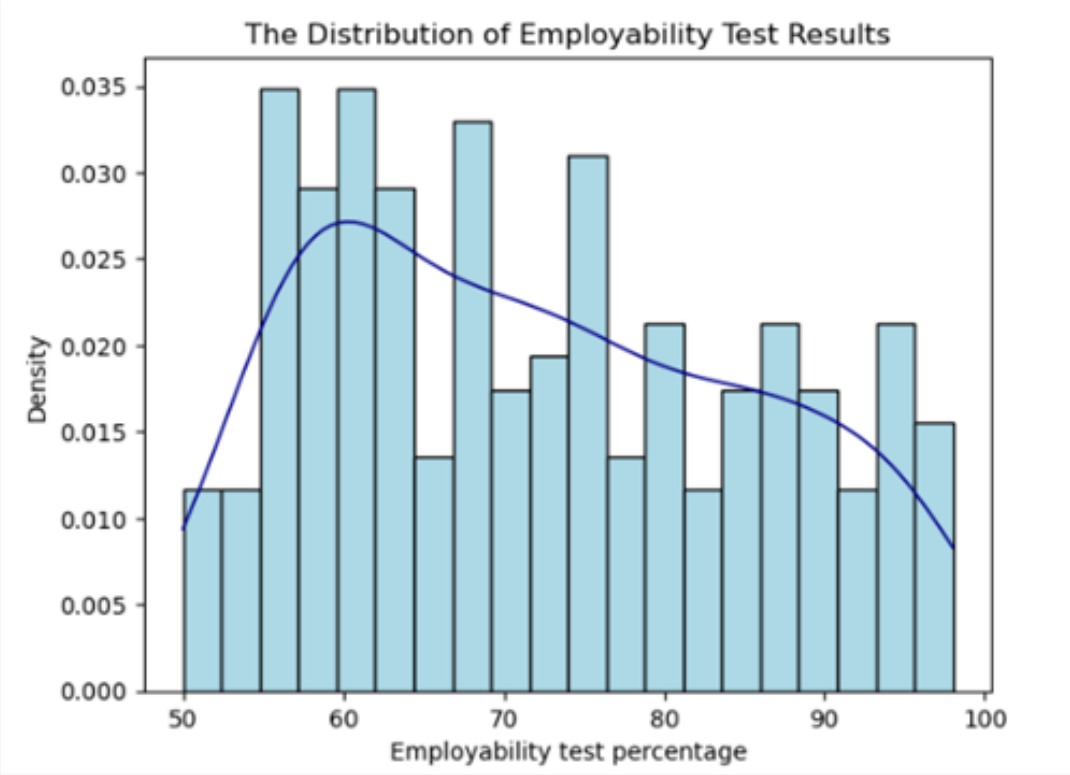
In addition, those have no experience have the same range as those having working experience (from about 53% to nearly 90%) when talking about secondary education percentage. The highest percentage in secondary education belonged to a work-inexperienced is nearly 90%, whilst the lowest rate belonged to work-experienced side.

PLOTS & EXPLANATION

03

Plot 3

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import gaussian_kde
from scipy.stats import gaussian_kde

a= pd.read_csv('Placement_Data_Full_Class.csv')
etest_p = a['etest_p']
kde = gaussian_kde(etest_p)
kde_x = np.linspace(etest_p.min(), etest_p.max(), 100)
kde_y = kde(kde_x)
plt.hist(etest_p, bins=20, density=True, color='lightblue', edgecolor='black')
plt.plot(kde_x, kde_y, color='darkblue', linewidth=1.2)
plt.title('The Distribution of Employability Test Results')
plt.xlabel('Employability test percentage')
plt.ylabel('Density')
plt.show()
```



PLOTS & EXPLANATION

Plot 3

The Employability test plays a crucial role in determining the employability of students after graduation. The histogram chart represents the distribution of the density of the employability test results conducted by the college. The chart clearly indicates that the highest percentage of students scored between 60% and 70% in the test. This indicates that these students have a higher chance of succeeding in the job market.

Furthermore, the lowest density of employability test scores was observed to be around 0.0005 of 100%, whereas the highest density of scores was over 0.03 for the 60%-70% test. Interestingly, students who have scored between 70%, 80%, and 90% Employability test have a density range between 0.025% to 0.005%, respectively.

In general, the test results indicate that students who scored below 60% have a reduced likelihood of succeeding in the job market. Therefore, it is crucial for students to aim for respectable scores in their Employability tests to secure better career prospects.

Through proper planning and practical training, students can improve their employability and their chances of entering the job market with a competitive advantage.

PLOTS & EXPLANATION

03

Plot 4

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
grouped = a.groupby(['hsc_s', 'workex', 'status']).size().reset_index(name='count')
placed_data = grouped[grouped['status'] == 'Placed']
not_placed_data = grouped[grouped['status'] == 'Not Placed']
placed_pivot = placed_data.pivot_table(index='hsc_s', columns='workex',
                                       values='count', fill_value=0)
not_placed_pivot = not_placed_data.pivot_table(index='hsc_s', columns='workex',
                                               values='count', fill_value=0)
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(8, 12), sharex=True)
placed_pivot.plot(kind='bar', stacked=True, ax=ax1)
ax1.legend(title='workex', loc='upper right')
ax1.tick_params(axis='x', rotation=0) # Rotate x-axis labels to horizontal
ax1.text(1.05, 0.5, 'Placed', transform=ax1.transAxes, ha='left', va='center', rotation=-90)
not_placed_pivot.plot(kind='bar', stacked=True, ax=ax2)
ax2.set_xlabel('Specialization in Higher Secondary Education')
ax2.set_ylabel('Number of candidates',fontsize=16)
ax2.yaxis.set_label_coords(-0.1, 1)
ax2.legend(title='workex', loc='upper right')
ax2.tick_params(axis='x', rotation=0)
ax2.text(1.05, 0.5, 'Not Placed', transform=ax2.transAxes, ha='left', va='center', rotation=-90)
fig.suptitle('The bar charts compare working experience and the status of placement in different fields')
plt.subplots_adjust(hspace=0.1)

plt.show()
```

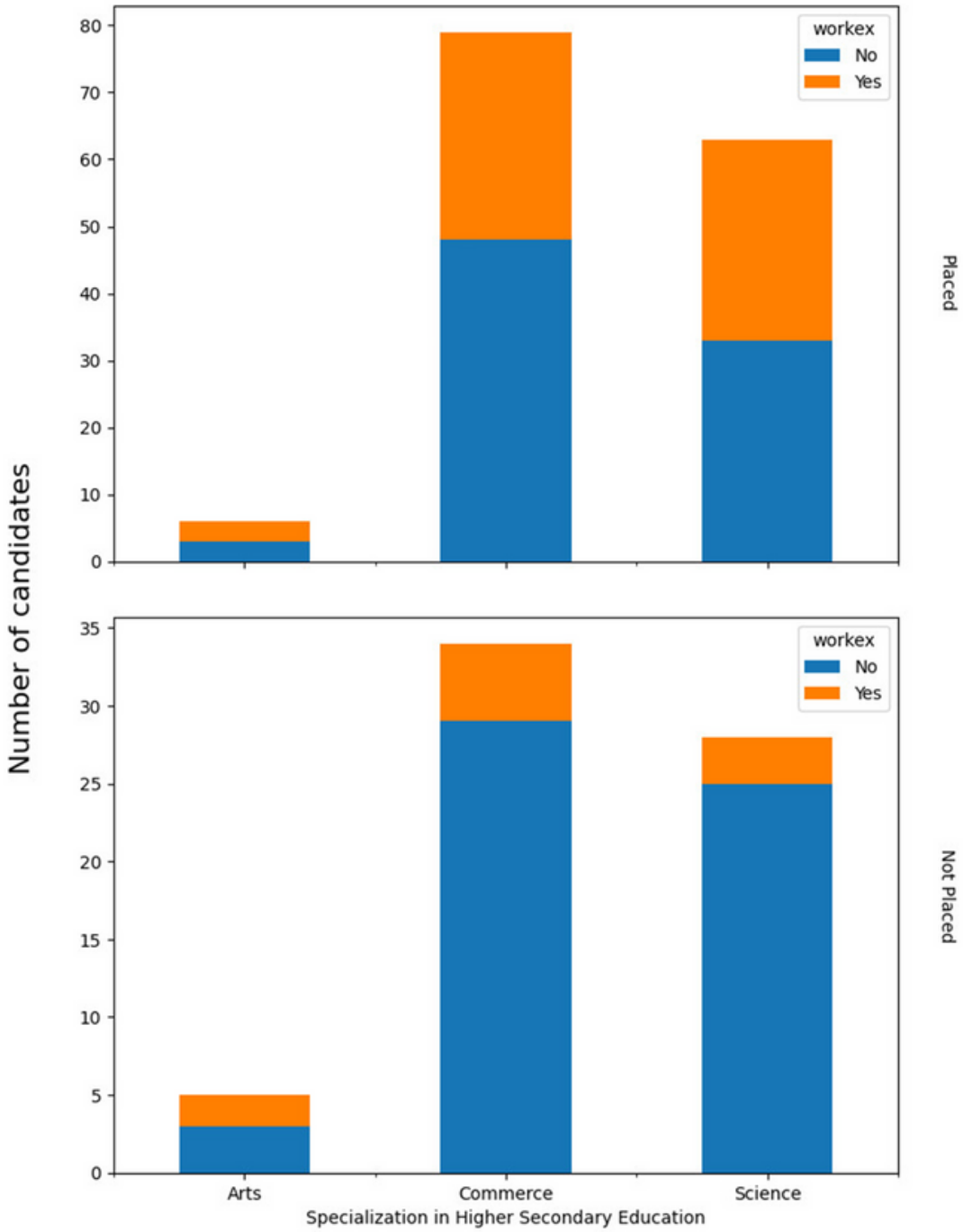
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 4

The bar chart compares working experience and the status of placement in different fields



PLOTS & EXPLANATION

03

Plot 4

The given chart shows the differences between candidates in three specializations in higher secondary education, including Arts, Commerce and Science. In general, there are more placed applicants than not placed candidates.

According to the first bar chart, the number of candidates not placed with working experience is below 5 in each specialization in higher secondary education. In the Commerce specialty, there are more than 30 candidates who were not selected, although there are fewer in the Science specialization. Overall, the majority of those who are not placed have never held a job.

Turning to the second chart, the majority is from Commerce specialization, with roughly 80 candidates. Arts has the lowest number of candidates at both sites (placed and not placed). Commerce and Science specializations have nearly the same number of candidates (about 30 for each one).

It can be seen that most candidates have no working experience, but on the placed site, this is not a big difference between yes and no working experience.

PLOTS & EXPLANATION

03

Plot 5

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
sns.scatterplot(data=a, x='hsc_p', y='degree_p', hue='status')
plt.title("The Chance of Having a Job")
plt.ylabel("Higher Secondary Education percentage-12th Grade")
plt.xlabel("Degree Percentage")
plt.show()
```

PLOTS & EXPLANATION

03

Plot 5

The given dot chart illustrates the relationship between three categories: the percentage of students who have a degree, the percentage of students who pass the 10th grade, and the status of work.

In general, we can see that the higher percentage of getting a higher score, the higher percentage of not being unemployed. As we can see on the chart, those people whose points are in the low zone of the chart are unemployed, as described by the red dots.

Furthermore, in the middle part of the chart, the situation is improved when they get the average points, and the chance of having a job is higher than in the low zone (about 50%).

Go to the top, and we can see that the blue dots outweigh the red ones because those people in this had the most points in total. We can only see a very few red dots on it. This chart mainly shows that if the students have a good score, they will have a better chance of getting a job in the future.

PLOTS & EXPLANATION

03

Plot 6

```
import matplotlib.pyplot as plt
import pandas as pd
a= pd.read_csv('Placement_Data_Full_Class.csv')
PD = a.groupby(['degree_t', 'gender']).size().reset_index(name='count')

total_counts = PD.groupby('degree_t')['count'].sum().reset_index()

PD['percentage'] = PD['count'] / PD.groupby('degree_t')['count'].transform('sum') * 100

fig, ax = plt.subplots(figsize=(8, 8))

sci_outer_color = '#F2BED1'
sci_inner_male_color = '#FDCEDF'
sci_inner_female_color = '#F8E8EE'

comm_outer_color = '#93C6E7'
comm_inner_male_color = '#AEE2FF'
comm_inner_female_color = '#B9F3FC'

others_outer_color = '#B1B2FF'
others_inner_male_color = '#AAC4FF'
others_inner_female_color = '#D2DAFF'

outer_labels = [f"{degree_t}\n{percentage:.1f}%" for degree_t, percentage in zip(total_counts['degree_t'],
                                    total_counts['count'] / total_counts['count'].sum() * 100)]
outer_wedges, outer_texts = ax.pie(total_counts['count'], labels=outer_labels, startangle=90,
                                   colors=[comm_outer_color, others_outer_color, sci_outer_color],
                                   radius=2.2, wedgeprops=dict(width=0.5, edgecolor='white'))
for text in outer_texts:text.set_horizontalalignment('center')
text.set_verticalalignment('center')
text.set_fontsize(10)
text.set_position((0.85 * text._x, 0.85 * text._y))
inner_labels = [f"{percentage:.1f}%" for percentage in PD['percentage']]
inner_gender_labels = PD['gender']
sci_inner_colors = [sci_inner_male_color, sci_inner_female_color]
comm_inner_colors = [comm_inner_male_color, comm_inner_female_color]
others_inner_colors = [others_inner_male_color, others_inner_female_color]
inner_colors = []
for row in PD.iterrows():degree_t = row[1]['degree_t']
gender = row[1]['gender']
if degree_t == 'Sci&Tech':inner_colors.append(sci_inner_colors[0 if gender == 'M' else 1])
elif degree_t == 'Comm&Mgmt':inner_colors.append(comm_inner_colors[0 if gender == 'M' else 1])
else:inner_colors.append(others_inner_colors[0 if gender == 'M' else 1])
inner_wedges, inner_texts = ax.pie(PD['count'], labels=inner_labels, startangle=90, colors=inner_colors,
                                   radius=1.7, wedgeprops=dict(width=0.5, edgecolor='white'))
```

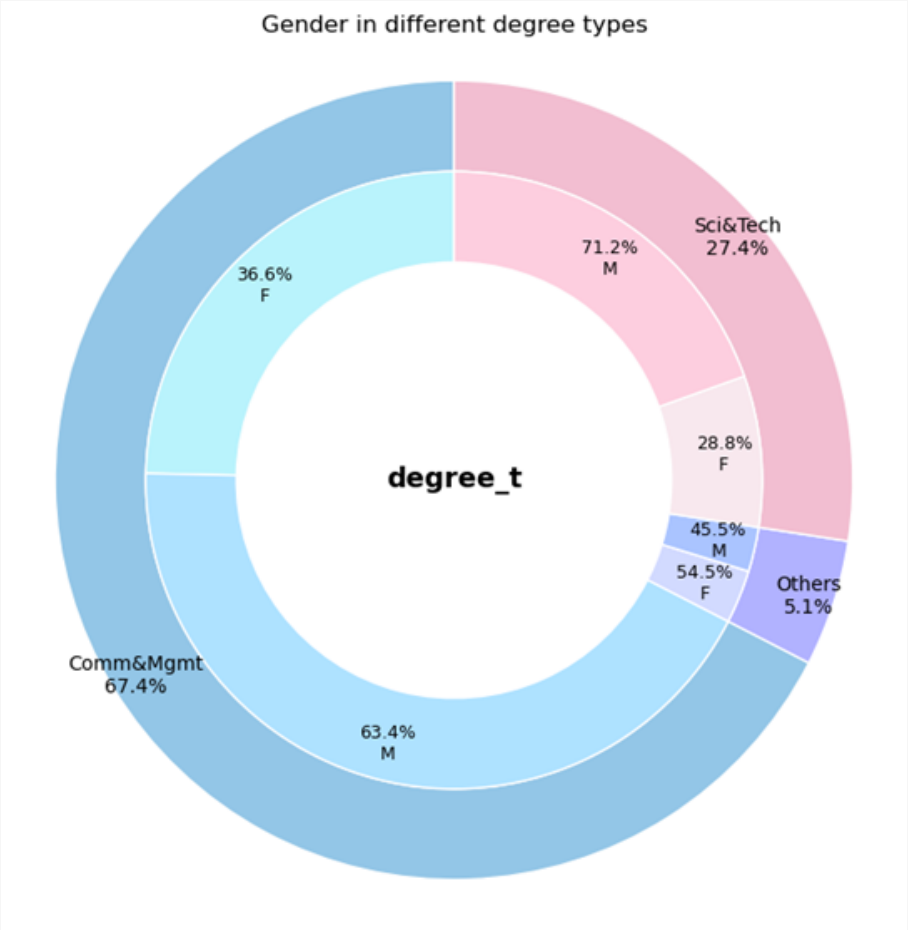
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 6

```
for text, label, gender_label in zip(inner_texts, inner_labels, inner_gender_labels):text.set_horizontalalignment('center')
text.set_verticalalignment('center')
text.set_rotation_mode('anchor')
text.set_rotation(0)
text.set_fontsize(9)
text.set_position((0.8 * text._x, 0.8 * text._y))
text.set_text(f"{label}\n{gender_label}")

center_circle = plt.Circle((0, 0), 0.5, color='white')
ax.add_artist(center_circle)
degree_t_label = 'degree_t'
ax.text(0, 0, degree_t_label, horizontalalignment='center', verticalalignment='center', fontsize=14, weight='bold')

ax.axis('equal')

ax.set_title("Gender in different degree types")

plt.show()
```



PLOTS & EXPLANATION

03

Plot 6

The pie chart shows information on different fields of degree education in universities for males and females. It can be clearly seen that the majority of degrees are in Communication and Management.

A high percentage of the males graduated with a Communication and Management degree, which makes up 63.4% of this field. The proportion of females in this type of degree is less than twice as high, with a percentage of 36.6%.

In terms of blue sites for the Science and Technology degree, which comprises 27%, the number of males is higher than that of females. In particular, males account for 71.2% and females for 28.8%.

For other types of degrees, the percentage of both genders seems not to be really different. Males consist of 45.5% in comparison with 54.5% of females. However, these types only make up a minority of candidates in total.

PLOTS & EXPLANATION

03

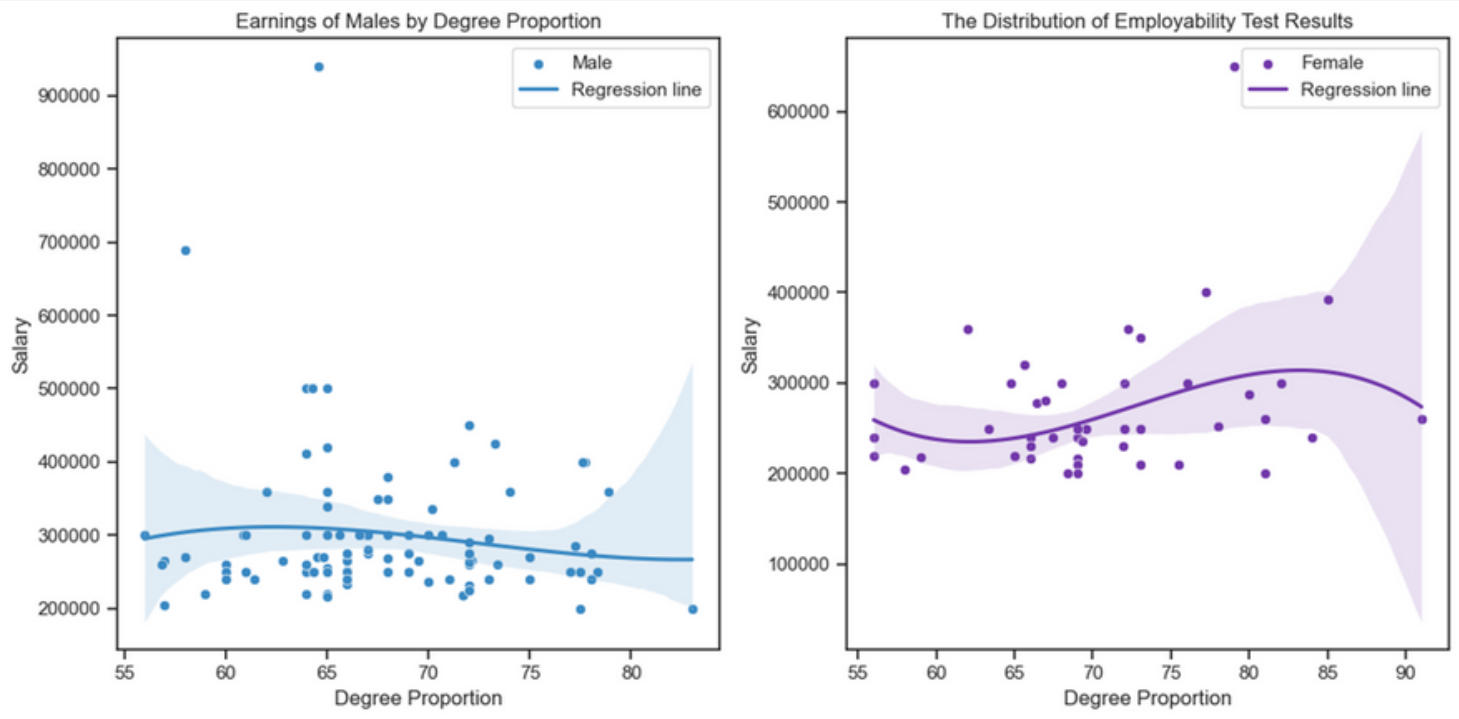
Plot 7

```
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
sns.set(style="ticks")

a= pd.read_csv('Placement_Data_Full_Class.csv')

male_color = "#3B8AC4"
female_color = "#7339AB"
male_data = a[a['gender'] == 'M']
female_data = a[a['gender'] == 'F']

fig, axes = plt.subplots(1, 2, figsize=(12, 6))

sns.scatterplot(data=male_data, x='degree_p', y='salary', color=male_color, label='Male', ax=axes[0])
sns.regplot(data=male_data, x='degree_p', y='salary', order=3, scatter=False, color=male_color,
            label='Regression line', ax=axes[0])

sns.scatterplot(data=female_data, x='degree_p', y='salary', color=female_color, label='Female', ax=axes[1])
sns.regplot(data=female_data, x='degree_p', y='salary', order=3, scatter=False, color=female_color,
            label='Regression line', ax=axes[1])

axes[0].set_title('Earnings of Males by Degree Proportion')
axes[0].set_xlabel('Degree Proportion')
axes[0].set_ylabel('Salary')
axes[0].legend()
axes[1].set_title('Earnings of Females by Degree Proportion')
axes[1].set_xlabel('Degree Proportion')
axes[1].set_ylabel('Salary')
axes[1].legend()
plt.tight_layout()

plt.show()
```

PLOTS & EXPLANATION

03

Plot 7



The graphical representation displays the earnings of boys and girls as per the proportion of degrees acquired in various fields. It is evident that students with degree percentages ranging from 60% to 80% earn more than $250,000 annually, depicting a positive correlation between degree percentage and income.

Further analysis reveals that individuals with degree percentages between 60% and 80% earn around $500,000 annually, indicating a strong financial incentive for individuals to pursue higher education. In contrast, students with graduation rates below 60% are not eligible for corporate job offers, making it extremely difficult for them to earn a high income.

PLOTS & EXPLANATION

03

Plot 7

From a broad perspective, these figures suggest that students who excel academically and attain higher proportions of degrees have greater opportunities to access high-paying jobs in the corporate world. As such, students who achieve above 60% of their degrees will have a greater chance of landing an esteemed corporate job and earning a large pay.

In conclusion, the graphical representation underscores the importance of education and its correlation with income. Individuals who prioritize their studies and aim for higher proportions of degrees have a heightened chance of securing well-paying jobs and living comfortable life.

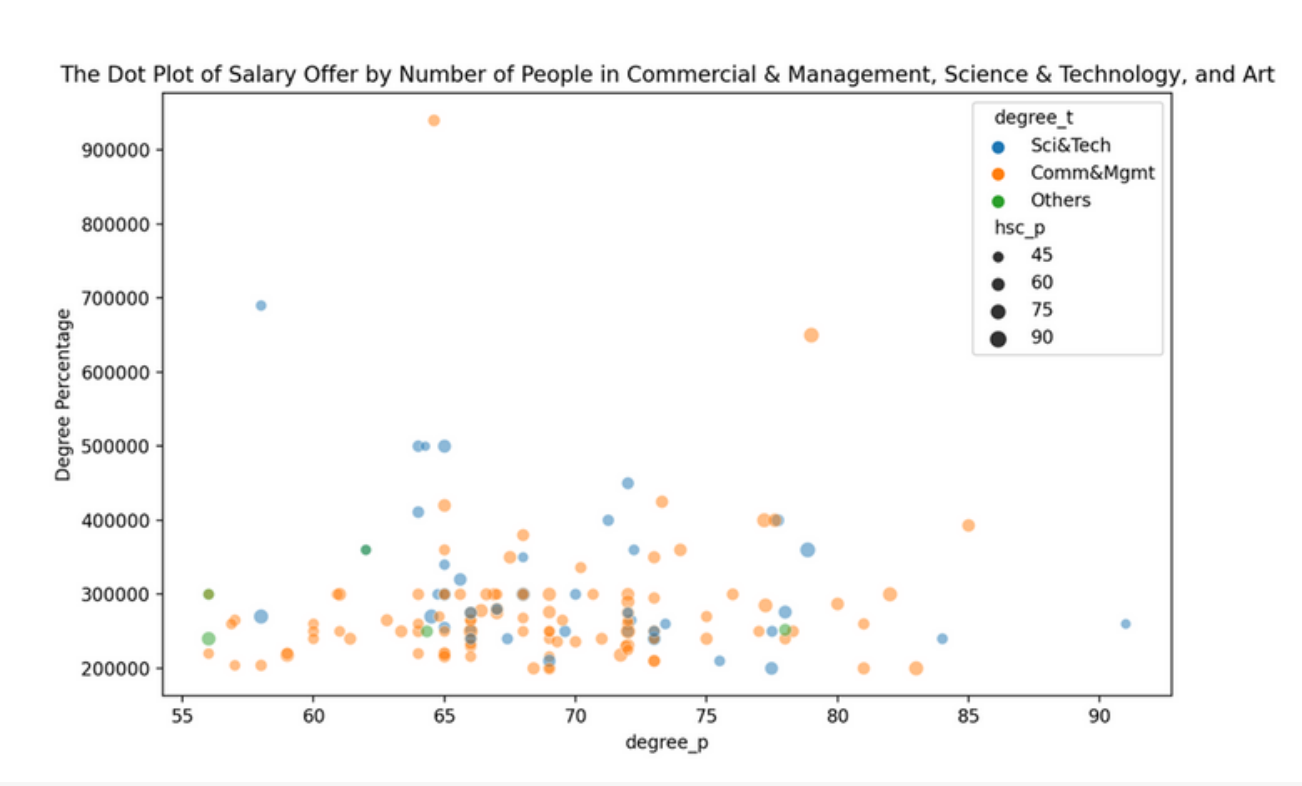
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 8

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(10, 6))
sns.scatterplot(data=a, x='degree_p', y='salary', hue='degree_t', size='hsc_p', alpha=0.5)
plt.title("The Dot Plot of Salary Offer by Number of People in Commercial & Management, Science & Technology, and Art")
plt.ylabel("Degree Percentage")
plt.show()
```



PLOTS & EXPLANATION

03

Plot 8

The graph illustrates the relationship between four categories, namely salary, degree percentages (degree_p), field of degree education (degree_t), and higher secondary education percentages (hsc_p), in a class at XYZ Campus.

At first glance, it is clear that most people who have one of three degrees have a salary between 200,000 USD and 400,000 USD. Additionally, there are few individuals who earn a high salary of more than 600,000 USD.

It could be noticed that people who work in commercial and management have degree percentages ranging from over 50% to nearly 85% mostly. Moreover, commercial and managerial positions have a wide range of salaries; it is clear to see that their salaries of them stretch from top to bottom. This field has a person who gets the highest income (over 950,000 USD) with only a 65% degree percentage and a 60% secondary school percentage.

Now, turning to the science and technology major, people working in this field have stable salaries, with some earning as much as $70,000 and some making as little as $200,000. Additionally, the person who gets the highest salary just takes over 55% in degree percentage and 60% in secondary school percentage. On the other hand, the one who gets over 90% in degree percentage earns a bit of money, which is less than $300,000.

Last but not least, the other majors, including arts, do not attract many students. However, there is one point on the graph that should be noted. The salary of those who work in other fields is quite low compared with the two mentioned specializations. Even the highest salary is less than $400,000. There are even people who achieve almost 80% in the degree, but the salary is just over $250,000.

PLOTS & EXPLANATION

03

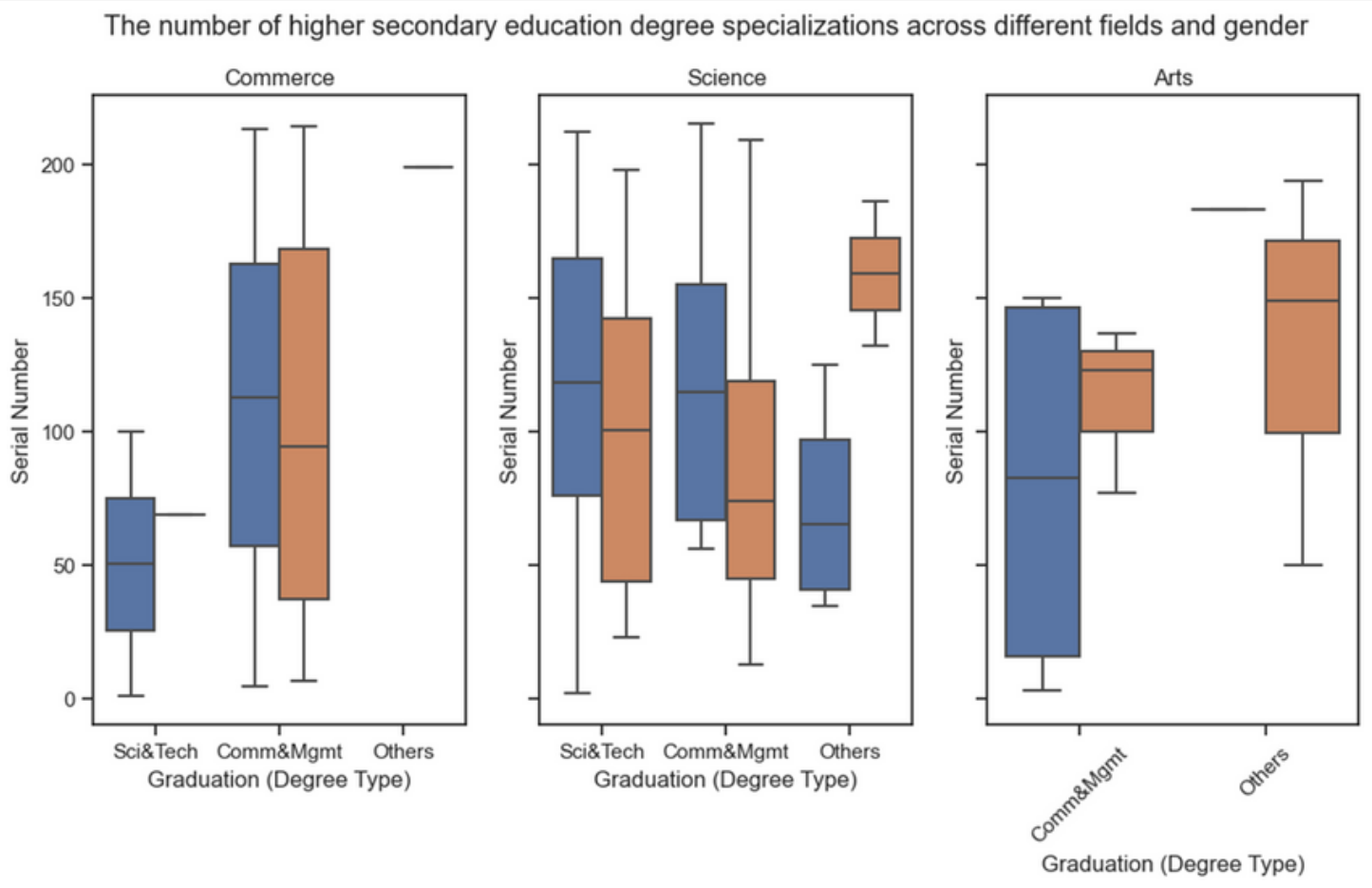
Plot 9

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
fig, axs = plt.subplots(1, len(a['hsc_s'].unique()), figsize=(12, 6), sharey=True)


for i, facet in enumerate(a['hsc_s'].unique()):

    facet_data = a[a['hsc_s'] == facet]
    ax = axs[i]

    sns.boxplot(data=facet_data, x='degree_t', y='sl_no', hue='gender', ax=ax)
    ax.set_xlabel("Graduation (Degree Type)")
    ax.set_ylabel("Serial Number")

    ax.get_legend().remove()
    ax.set_title(facet)
    plt.xticks(rotation=45)


fig.suptitle("The number of higher secondary education degree specializations across different fields and gender")

handles, labels = axs[-1].get_legend_handles_labels()
fig.legend(handles, labels, loc='lower right')

plt.tight_layout()

plt.show()
```

PLOTS & EXPLANATION

03

Plot 9



PLOTS & EXPLANATION

03

Plot 9

The boxplot graph provides an overview of the number of higher secondary education degree specializations across different fields and gender. The data indicates that comm&mgmt is the most commonly pursued degree, with more than 50 serial number students enrolled in this field. On the other hand, others fields have a significantly lower number of enrollments.

When viewed as a whole, it appears that the preference for degree specializations is significantly influenced by the type of discipline. For instance, the higher secondary arts education has a higher proportion of non-science and technology degrees as compared to others, which are mainly focused on science and technology.

Moreover, the boxplot graph also highlights gender-based differences in the choice of higher secondary education degrees. Female students tend to lean towards commerce as compared to their male counterparts who prefer to pursue degrees in science and technology. This could be attributed to gender-based stereotypes that often associate women with management and administrative roles and men with research and technical positions.

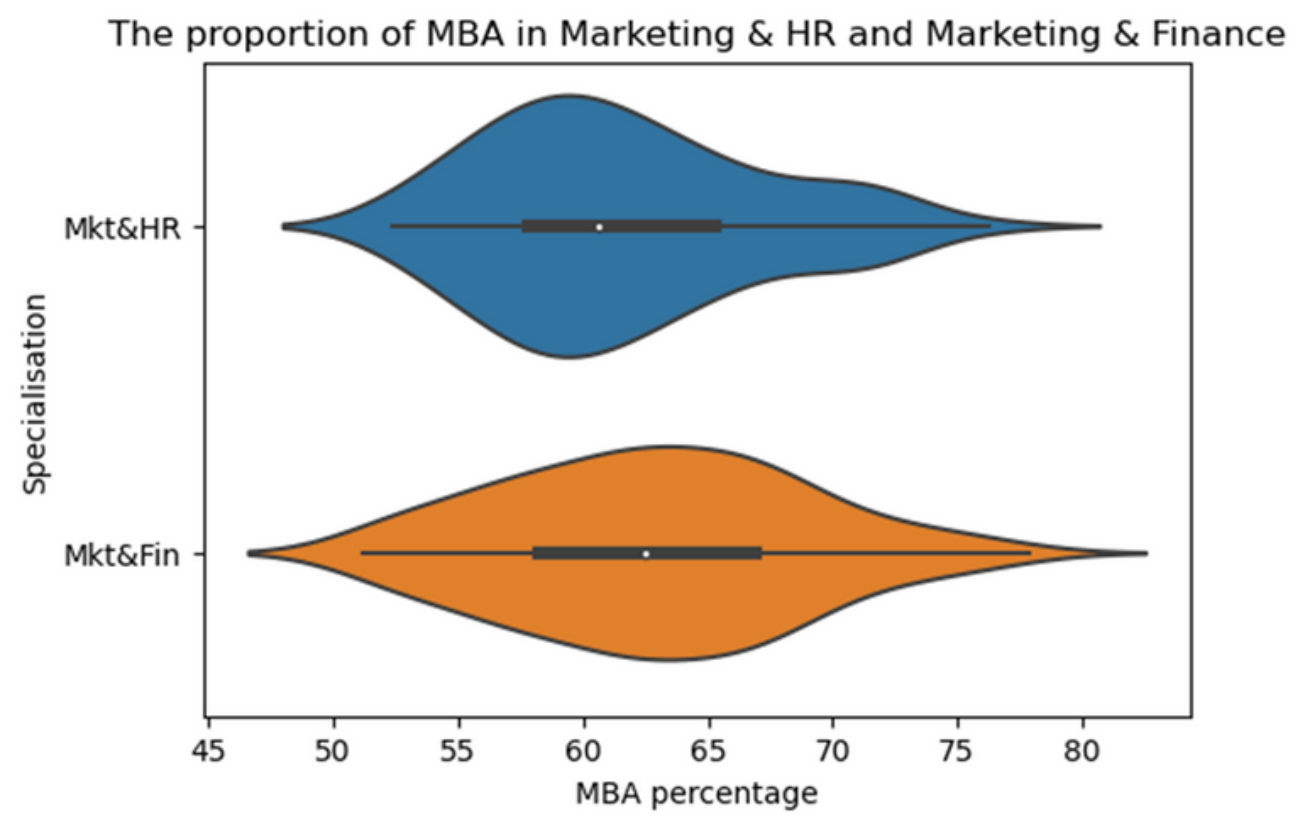
Overall, the boxplot graph offers valuable insights into the enrollment trends of higher secondary education degree specializations across various fields and gender. It underscores the importance of considering the influence of societal and cultural factors on academic choices, particularly among female students.

PLOTS & EXPLANATION

03

Plot 10

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

%matplotlib inline
a = pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(6,4), dpi=100)
sns.violinplot(data = a, x ='mba_p', y = 'specialisation')
plt.title("The proportion of MBA in Marketing & HR and Marketing & Finance")
plt.ylabel("Specialisation")
plt.xlabel("MBA percentage")
plt.show()
```



PLOTS & EXPLANATION

03

Plot 10

As the violin chart provided shows, it is evident that the proportion of MBA students in the Marketing & Finance specialization is greater than that in the Marketing & HR specialization. It can be observed that a majority of the Marketing & Finance students have scores above 60%, while most of the Marketing & HR students have scores between 55 and 65.

Upon further analysis, it was found that fewer Marketing & Finance students had scores below 60% as compared to the Marketing & HR students. Additionally, it was discovered that there were more students in the Marketing & Finance specialization than those in the Marketing & HR specialization, with over 60% more students.

Moreover, it has been observed that a higher percentage of students in the Marketing & Finance specialization have an MBA degree as compared to those in the Marketing & HR specialization. In fact, the number of students in the Marketing & Finance specialization with an MBA degree outweighs those in the Marketing & HR specialization by more than 70%.

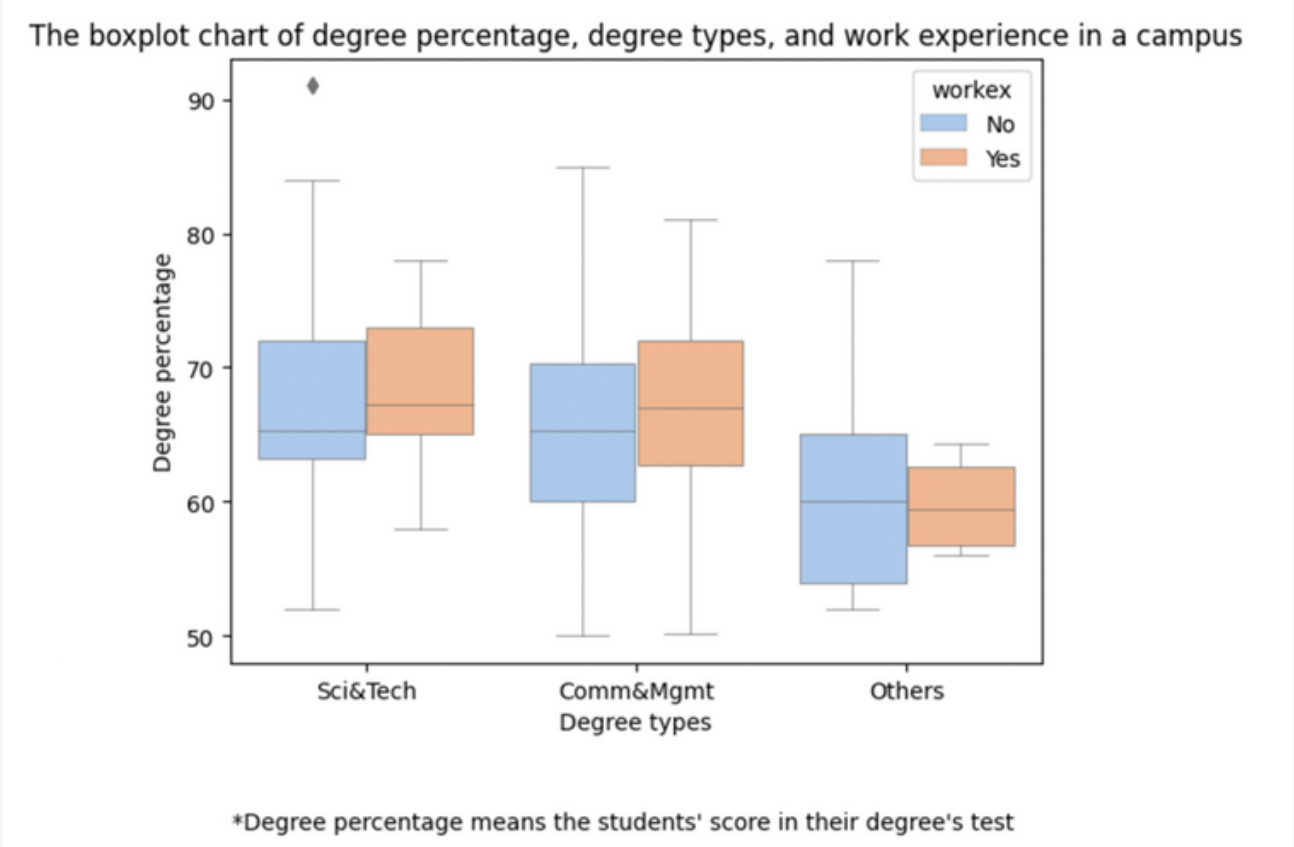
Overall, these findings indicate that the Marketing & Finance specialization is more popular among MBA students as compared to the Marketing & HR specialization.

PLOTS & EXPLANATION

03

Plot 11

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

a = pd.read_csv('Placement_Data_Full_Class.csv')
sns.boxplot(x='degree_t', y='degree_p', hue='workex', data=a, width=0.8, linewidth=0.5, palette='pastel')
plt.title("The boxplot chart of degree percentage, degree types, and work experience in a campus")
plt.xlabel("Degree types")
plt.ylabel("Degree percentage")
plt.figtext(0.5, -0.1, "*Degree percentage means the students' score in their degree's test", ha='center')
plt.show()
```



PLOTS & EXPLANATION

03

Plot 11

The provided boxplot chart demonstrate the connection of degree percentage, degree types (Communication and Management, Science and Technology and others) and work experience of a campus.

As we can see from the chart, the percentage of degree of those studying Communication and Management, Science and Technology almost has the same interquartile range. In contrast, the degree percentage of other students has lower interquartile range.

Starting with those studying Communication and Management, people who have work experience (from about 63% to approximately 73%) has the higher range in total than the other (from 60% to above 70%). Therefore, the average rate of those working for the first time and having worked before is about 65% and nearly 67%, in respectively. In addition, those have no experience are the main students in the campus.

In terms of other degrees, the percentage of those having no work experience has the widest range than any categories (from under 55% to exactly 65%). On the other hand, the percentage of people having worked before has smaller range and its average is nearly the same as the other (about 59% and 60% in respectively).

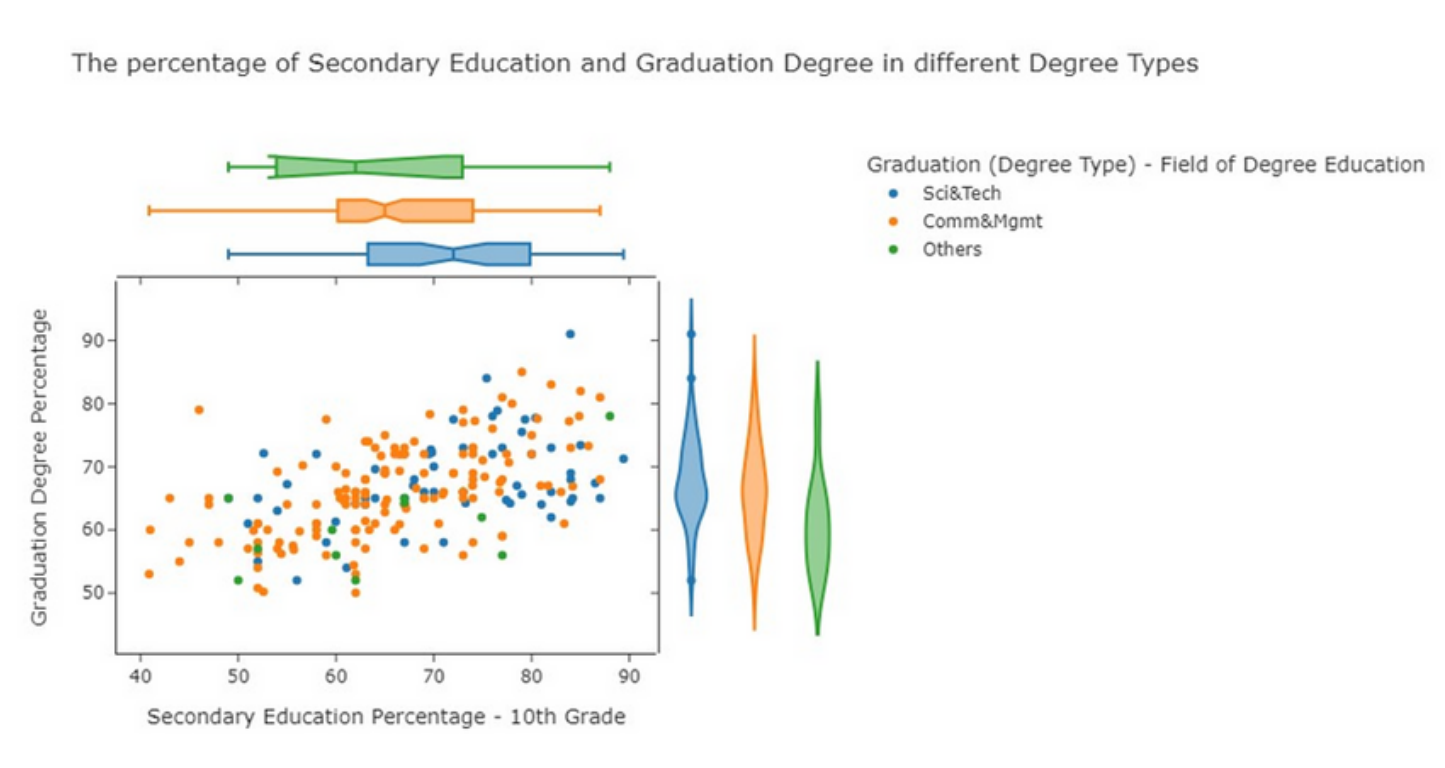
Turning to Science and Technology, it has the same average rate in both sides as those studying Communication and Management. Additionally, its interquartile range is also the same as those studying Communication and Management.

PLOTS & EXPLANATION

03

Plot 12

```
import pandas as pd
import plotly.express as px

a = pd.read_csv('Placement_Data_Full_Class.csv')
fig = px.scatter(a, x="ssc_p", y="degree_p", color="degree_t", marginal_y="violin",
                 marginal_x="box", template="simple_white",
labels={'ssc_p': 'Secondary Education Percentage - 10th Grade',
        'degree_p': 'Graduation Degree Percentage',
        'degree_t': 'Graduation (Degree Type) - Field of Degree Education'},
title='The percentage of Secondary Education and Graduation Degree in different Degree Types')

fig.show()
```



PLOTS & EXPLANATION

03

Plot 12

The visual representations in the form of point, box, and violin charts offer valuable insights into the educational landscape of graduating students. The point chart effectively captures the density distribution, aligning with the patterns observed in industries, while also providing information on the ratio of secondary education percentages among students. In the realm of higher education, an analysis of graduating students' degree choices reveals intriguing trends. The combined numbers of students pursuing degrees in Science & Technology (Sci&Tech) and Commerce & Management (Comm&Mgmt) fields comprise a significant portion of the total graduating cohort. However, it is important to note that students from other fields also contribute, albeit to a lesser extent.

Delving deeper into the data, an interesting correlation emerges between degree choices and secondary education performance. Notably, a larger number of secondary students, scoring between 60% and 75%, opt for degrees in Comm&Mgmt. Within this group, degree percentages range from 55% to 75%. Conversely, students graduating in other fields tend to possess an average degree percentage of approximately 60%, alongside secondary education scores below 80%.

To summarize, while Sci&Tech and Comm&Mgmt fields dominate a significant portion of graduating students, there is also representation from other disciplines. The relationship between degree choices, secondary education performance, and the varying distributions across fields highlight the diverse academic landscape of graduating students.

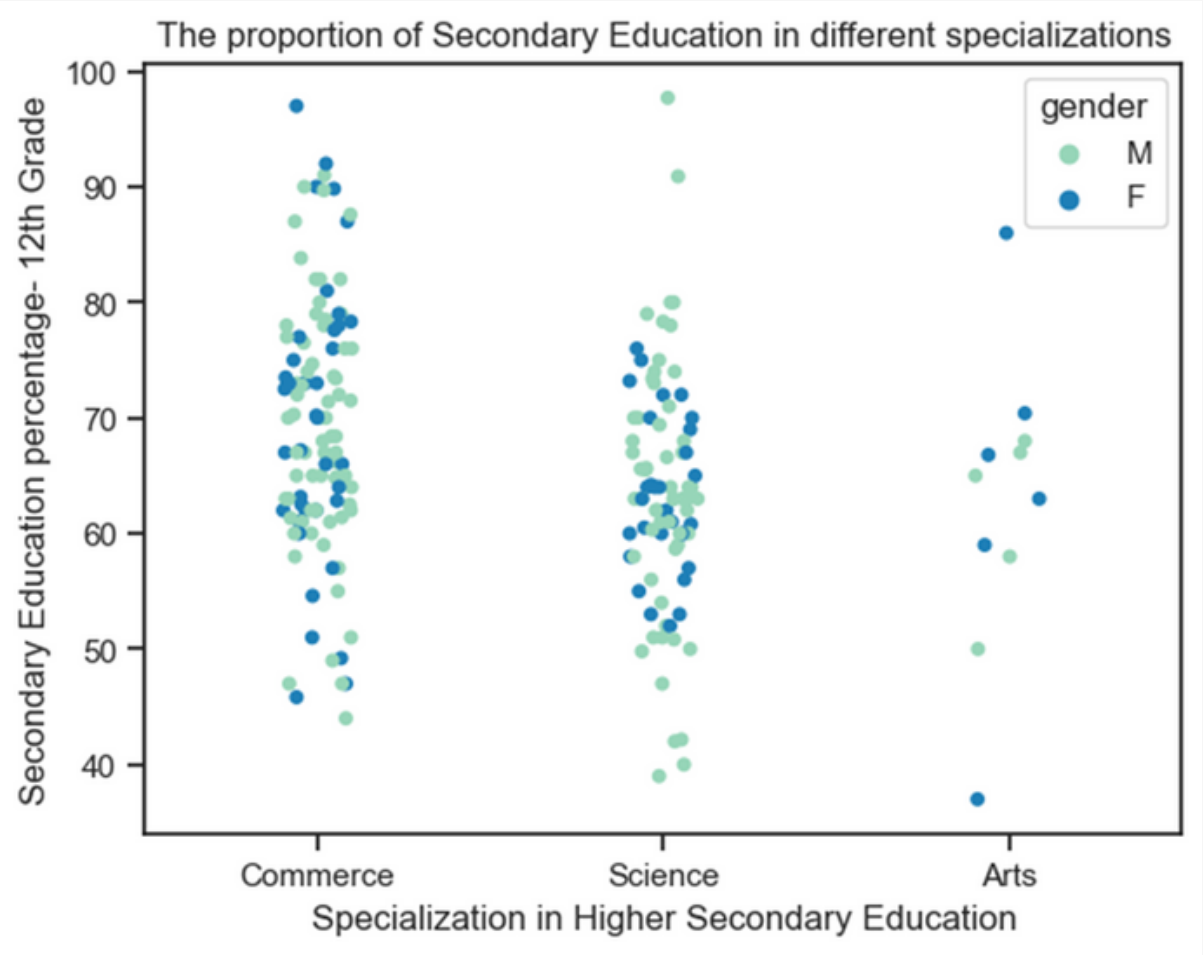
Vietnamese-German University

PLOTS & EXPLANATION

03

Plot 13

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

a = pd.read_csv('Placement_Data_Full_Class.csv')
sns.stripplot(x ='hsc_s', y = 'hsc_p', data = a, hue = 'gender', palette = 'YlGnBu')
plt.title("The proportion of Secondary Education in different specializations")
plt.ylabel("Secondary Education percentage- 12th Grade")
plt.xlabel("Specialization in Higher Secondary Education")
plt.show()
```



PLOTS & EXPLANATION

03

Plot 13

Commerce and Science are two prominent specializations in Higher Secondary Education, attracting a substantial number of students. However, there is a noticeable difference in the gender distribution within these fields, with a higher number of male students compared to female students.

Commerce programs have gained popularity among students who have achieved approximately 70% scores in their 12th-grade examinations. This indicates a growing interest in business and finance-related subjects among the youth. Similarly, Science programs have also witnessed significant enrollment, with students have approximately 60% choosing Science as their specialization. In contrast, the Arts specialization has shown comparatively lower popularity, as evidenced by the number of Secondary students (only 11) achieving over 60% in their Secondary Education in this field.

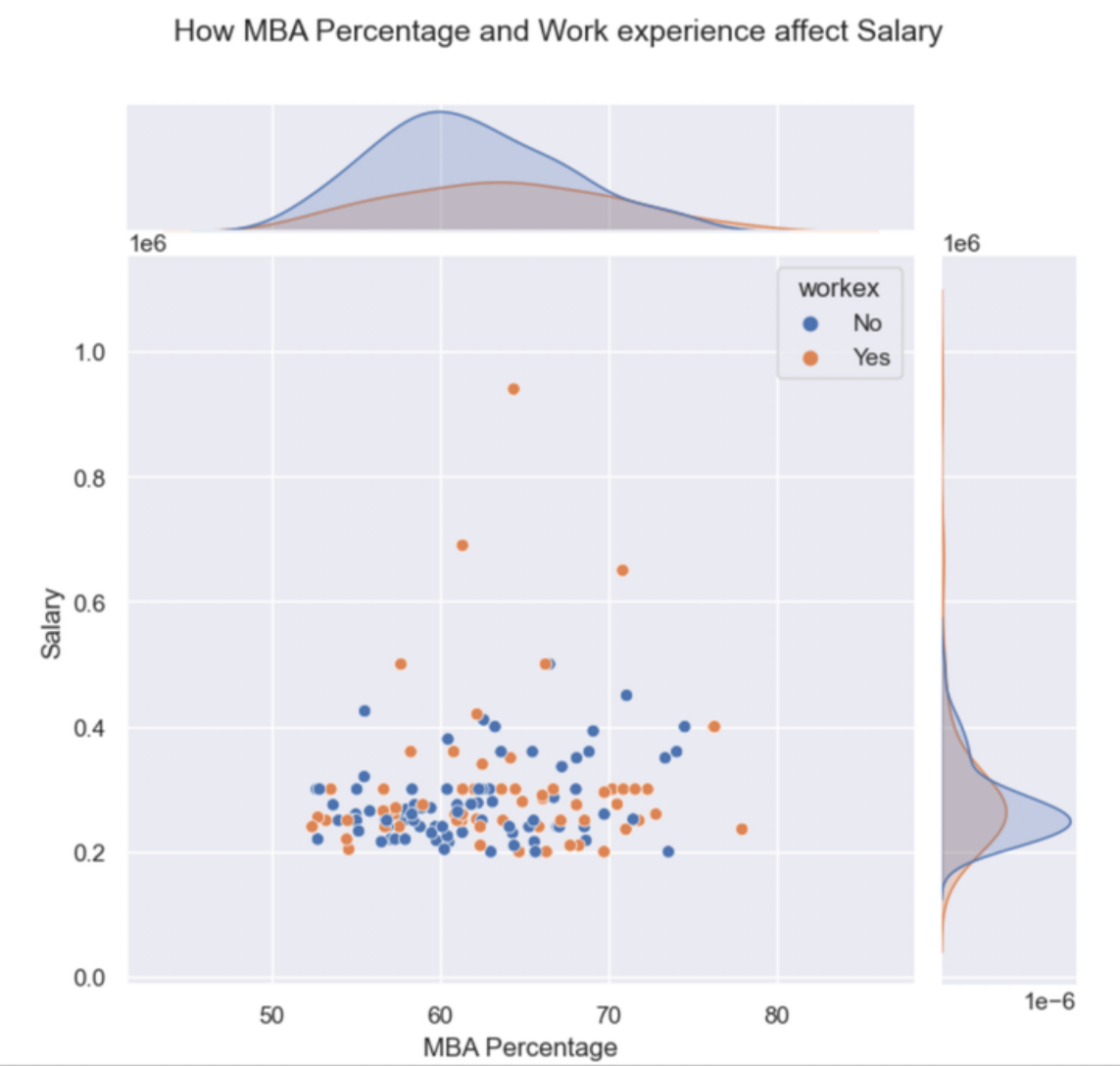
Overall, the distribution of specializations in Higher Secondary Education highlights the diverse preferences and career aspirations of students. Commerce and Science emerge as dominant fields, with a higher number of male students opting for these specializations. The balanced gender representation across specializations emphasizes the importance of providing equal opportunities and support to both male and female students in pursuing their chosen fields of study.

PLOTS & EXPLANATION

03

Plot 14

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

a=pd.read_csv('Placement_Data_Full_Class.csv')
sns.set(style="darkgrid")
g=sns.jointplot(x='mba_p', y='salary', data=a, hue='workex', height=7)
g.set_axis_labels("MBA Percentage", "Salary")
g.fig.suptitle("How MBA Percentage and Work experience affect Salary", fontsize=14)
g.fig.subplots_adjust(top=0.9)
plt.show()
```



PLOTS & EXPLANATION

03

Plot 14

The given chart reveals the connection between the salary of a year offered to students by companies, the students' MBA result and their work experience. Whilst the dot chart gives a detailed look at variables, two area charts aside provide the statistics in a more general way.

From the chart, it can be inferred that most students have their score in a range between nearly 50% and about 80%. In addition, the salary offered to students by many companies is mainly from nearly 200000 dollars to about 500000 dollars.

Beginning with the MBA result, the scores of the 2 majors vary from just over 50% to nearly 80%. However, there are a few students who are experienced in working and gain the highest scores (nearly 80%). Therefore, those who are inexperienced in working tend to normally get their score around 60%. Whilst the others are likely to get their MBA result between 60% and 70%.

Turning to the salary in a year offered to students by companies, those who are newcomers to the job market have a higher chance to get paid about 250000 dollars. Meanwhile, many students attending the job market for a long time are offered a higher salary. Some of them can be paid nearly 1000000 dollars a year.

APPENDIX: CODES FOR 14 PLOTS

Plot 1:

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
a= pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(8, 6))
sns.countplot(data=a, x='hsc_s', hue='gender')
plt.title("The Number of Men and Women Working in Arts, Commerce, and Science")
plt.xlabel("Field")
plt.show()
```

Plot 2:

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
a= pd.read_csv('Placement_Data_Full_Class.csv')
sns.scatterplot(data=a, y='salary',x='ssc_p',hue='workex')
plt.title("The dot chart of salary offered by companies to candidates, second education
percentage and work experience")
plt.xlabel("Secondary education percentage")
plt.ylabel("Salary")
plt.legend(title='Work experience')
plt.show()
```

Plot 3:

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import gaussian_kde
from scipy.stats import gaussian_kde

a= pd.read_csv('Placement_Data_Full_Class.csv')
etest_p = a['etest_p']
kde = gaussian_kde(etest_p)
kde_x = np.linspace(etest_p.min(), etest_p.max(), 100)
kde_y = kde(kde_x)
plt.hist(etest_p, bins=20, density=True, color='lightblue', edgecolor='black')
```

APPENDIX: CODES FOR 14 PLOTS

```
plt.plot(kde_x, kde_y, color='darkblue', linewidth=1.2)
plt.title('The Distribution of Employability Test Results')
plt.xlabel('Employability test percentage')
plt.ylabel('Density')
plt.show()
```

Plot 4:

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
grouped = a.groupby(['hsc_s', 'workex', 'status']).size().reset_index(name='count')
placed_data = grouped[grouped['status'] == 'Placed']
not_placed_data = grouped[grouped['status'] == 'Not Placed']
placed_pivot = placed_data.pivot_table(index='hsc_s', columns='workex', values='count',
fill_value=0)
not_placed_pivot = not_placed_data.pivot_table(index='hsc_s', columns='workex',
values='count', fill_value=0)
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(8, 12), sharex=True)
placed_pivot.plot(kind='bar', stacked=True, ax=ax1)
ax1.legend(title='workex', loc='upper right')
ax1.tick_params(axis='x', rotation=0) # Rotate x-axis labels to horizontal
ax1.text(1.05, 0.5, 'Placed', transform=ax1.transAxes, ha='left', va='center', rotation=-90)
not_placed_pivot.plot(kind='bar', stacked=True, ax=ax2)
ax2.set_xlabel('Specialization in Higher Secondary Education')
ax2.set_ylabel('Number of candidates',fontsize=16)
ax2.yaxis.set_label_coords(-0.1, 1)
ax2.legend(title='workex', loc='upper right')
ax2.tick_params(axis='x', rotation=0)
ax2.text(1.05, 0.5, 'Not Placed', transform=ax2.transAxes, ha='left', va='center', rotation=-90)
fig.suptitle('The bar charts compare working experience and the status of placement in
different fields')
plt.subplots_adjust(hspace=0.1)

plt.show()
```

APPENDIX: CODES FOR 14 PLOTS

Plot 5:

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
sns.scatterplot(data=a, x='hsc_p', y='degree_p', hue='status')
plt.title("The Chance of Having a Job")
plt.ylabel("Higher Secondary Education percentage-12th Grade")
plt.xlabel("Degree Percentage")
plt.show()
```

Plot 6:

```
import matplotlib.pyplot as plt
import pandas as pd
a= pd.read_csv('Placement_Data_Full_Class.csv')
PD = a.groupby(['degree_t', 'gender']).size().reset_index(name='count')

total_counts = PD.groupby('degree_t')['count'].sum().reset_index()

PD['percentage'] = PD['count'] / PD.groupby('degree_t')['count'].transform('sum') * 100

fig, ax = plt.subplots(figsize=(8, 8))

sci_outer_color = '#F2BED1'
sci_inner_male_color = '#FDCEDF'
sci_inner_female_color = '#F8E8EE'

comm_outer_color = '#93C6E7'
comm_inner_male_color = '#AEE2FF'
comm_inner_female_color = '#B9F3FC'

others_outer_color = '#B1B2FF'
others_inner_male_color = '#AAC4FF'
others_inner_female_color = '#D2DAFF'
```

APPENDIX: CODES FOR 14 PLOTS

```
outer_labels = [f"{degree_t}\n{percentage:.1f}%" for degree_t, percentage in
zip(total_counts['degree_t'],
total_counts['count'] / total_counts['count'].sum() * 100)]
outer_wedges, outer_texts = ax.pie(total_counts['count'], labels=outer_labels,
startangle=90,
colors=[comm_outer_color, others_outer_color, sci_outer_color],
                        radius=2.2, wedgeprops=dict(width=0.5, edgecolor='white'))

for text in outer_texts:text.set_horizontalalignment('center')
text.set_verticalalignment('center')
text.set_fontsize(10)
text.set_position((0.85 * text._x, 0.85 * text._y))
inner_labels = [f"{percentage:.1f}%" for percentage in PD['percentage']]
inner_gender_labels = PD['gender']
sci_inner_colors = [sci_inner_male_color, sci_inner_female_color]
comm_inner_colors = [comm_inner_male_color, comm_inner_female_color]
others_inner_colors = [others_inner_male_color, others_inner_female_color]
inner_colors = []
for row in PD.iterrows():degree_t = row[1]['degree_t']
gender = row[1]['gender']
if degree_t == 'Sci&Tech':inner_colors.append(sci_inner_colors[0 if gender == 'M' else 1])
elif degree_t == 'Comm&Mgmt':inner_colors.append(comm_inner_colors[0 if gender ==
'M' else 1])
else:inner_colors.append(others_inner_colors[0 if gender == 'M' else 1])
inner_wedges, inner_texts = ax.pie(PD['count'], labels=inner_labels, startangle=90,
colors=inner_colors,
                        radius=1.7, wedgeprops=dict(width=0.5, edgecolor='white'))

for text, label, gender_label in zip(inner_texts, inner_labels,
inner_gender_labels):text.set_horizontalalignment('center')
text.set_verticalalignment('center')
text.set_rotation_mode('anchor')
text.set_rotation(0)
text.set_fontsize(9)
text.set_position((0.8 * text._x, 0.8 * text._y))
text.set_text(f"{label}\n{gender_label}")
```

APPENDIX: CODES FOR 14 PLOTS

```
center_circle = plt.Circle((0, 0), 0.5, color='white')
ax.add_artist(center_circle)
degree_t_label = 'degree_t'
ax.text(0, 0, degree_t_label, horizontalalignment='center', verticalalignment='center',
fontsize=14, weight='bold')

ax.axis('equal')

ax.set_title("Gender in different degree types")

plt.show()
```

Plot 7:

```
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
sns.set(style="ticks")

a= pd.read_csv('Placement_Data_Full_Class.csv')

male_color = "#3B8AC4"
female_color = "#7339AB"
male_data = a[a['gender'] == 'M']
female_data = a[a['gender'] == 'F']

fig, axes = plt.subplots(1, 2, figsize=(12, 6))

sns.scatterplot(data=male_data, x='degree_p', y='salary', color=male_color, label='Male',
ax=axes[0])
sns.regplot(data=male_data, x='degree_p', y='salary', order=3, scatter=False,
color=male_color,
label='Regression line', ax=axes[0])
sns.scatterplot(data=female_data, x='degree_p', y='salary', color=female_color,
label='Female', ax=axes[1])
sns.regplot(data=female_data, x='degree_p', y='salary', order=3, scatter=False,
color=female_color,
label='Regression line', ax=axes[1])
```

APPENDIX: CODES FOR 14 PLOTS

```
axes[0].set_title('Earnings of Males by Degree Proportion')
axes[0].set_xlabel('Degree Proportion')
axes[0].set_ylabel('Salary')
axes[0].legend()
axes[1].set_title('Earnings of Females by Degree Proportion')
axes[1].set_xlabel('Degree Proportion')
axes[1].set_ylabel('Salary')
axes[1].legend()
plt.tight_layout()

plt.show()
```

Plot 8:

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

a= pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(10, 6))
sns.scatterplot(data=a, x='degree_p', y='salary', hue='degree_t', size='hsc_p', alpha=0.5)
plt.title("The Dot Plot of Salary Offer by Number of People in Commercial & Management,
Science & Technology, and Art")
plt.ylabel("Degree Percentage")
plt.show()
```

Plot 9:

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
a= pd.read_csv('Placement_Data_Full_Class.csv')
fig, axs = plt.subplots(1, len(a['hsc_s'].unique()), figsize=(12, 6), sharey=True)

for i, facet in enumerate(a['hsc_s'].unique()):

  facet_data = a[a['hsc_s'] == facet]
  ax = axs[i]
```

APPENDIX: CODES FOR 14 PLOTS

```
sns.boxplot(data=facet_data, x='degree_t', y='sl_no', hue='gender', ax=ax)
ax.set_xlabel("Graduation (Degree Type)")
  ax.set_ylabel("Serial Number")

ax.get_legend().remove()
  ax.set_title(facet)
plt.xticks(rotation=45)

fig.suptitle("The number of higher secondary education degree specializations across
different fields and gender")

handles, labels = axs[-1].get_legend_handles_labels()
fig.legend(handles, labels, loc='lower right')

plt.tight_layout()

plt.show()
```

Plot 10:

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

%matplotlib inline
a = pd.read_csv('Placement_Data_Full_Class.csv')
plt.figure(figsize=(6,4), dpi=100)
sns.violinplot(data = a, x ='mba_p', y = 'specialisation')
plt.title("The proportion of MBA in Marketing & HR and Marketing & Finance")
plt.ylabel("Specialisation")
plt.xlabel("MBA percentage")
plt.show()
```

Plot 11:

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

APPENDIX: CODES FOR 14 PLOTS

```
a = pd.read_csv('Placement_Data_Full_Class.csv')
sns.boxplot(x='degree_t', y='degree_p', hue='workex', data=a, width=0.8, linewidth=0.5,
palette='pastel')
plt.title("The boxplot chart of degree percentage, degree types, and work experience in a
campus")
plt.xlabel("Degree types")
plt.ylabel("Degree percentage")
plt.figtext(0.5, -0.1, "*Degree percentage means the students' score in their degree's test",
ha='center')
plt.show()
```

Plot 12:

```
import pandas as pd
import plotly.express as px

a = pd.read_csv('Placement_Data_Full_Class.csv')
fig = px.scatter(a, x="ssc_p", y="degree_p", color="degree_t", marginal_y="violin",
           marginal_x="box", template="simple_white",
labels={'ssc_p': 'Secondary Education Percentage - 10th Grade',
     'degree_p': 'Graduation Degree Percentage',
     'degree_t': 'Graduation (Degree Type) - Field of Degree Education'},
title='The percentage of Secondary Education and Graduation Degree in different Degree
Types')

fig.show()
```

Plot 13:

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

a = pd.read_csv('Placement_Data_Full_Class.csv')
sns.stripplot(x ='hsc_s', y = 'hsc_p', data = a, hue = 'gender', palette = 'YlGnBu')
plt.title("The proportion of Secondary Education in different specializations")
plt.ylabel("Secondary Education percentage- 12th Grade")
plt.xlabel("Specialization in Higher Secondary Education")
plt.show()
```

APPENDIX: CODES FOR 14 PLOTS

Plot 14:

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

a=pd.read_csv('Placement_Data_Full_Class.csv')
sns.set(style="darkgrid")
g=sns.jointplot(x='mba_p', y='salary', data=a, hue='workex', height=7)
g.set_axis_labels("MBA Percentage", "Salary")
g.fig.suptitle("How MBA Percentage and Work experience affect Salary", fontsize=14)
g.fig.subplots_adjust(top=0.9)
plt.show()
```