

DATA VISUALIZATION USING SEABORN

Placement Dataset Analysis

This document provides insights derived from a placement dataset using Seaborn visualizations. The following plots are utilized to analyze and explain the data:

- ❖ **Violinplot:** For understanding salary distributions across genders in different domains.
- ❖ **Boxplot:** For comparing score ranges and identifying key differences between groups.
- ❖ **Pointplot:** For visualizing performance comparisons across genders and boards.
- ❖ **Catplot:** For exploring categorical data trends in the dataset.

❖ Why we use Swarmplot?

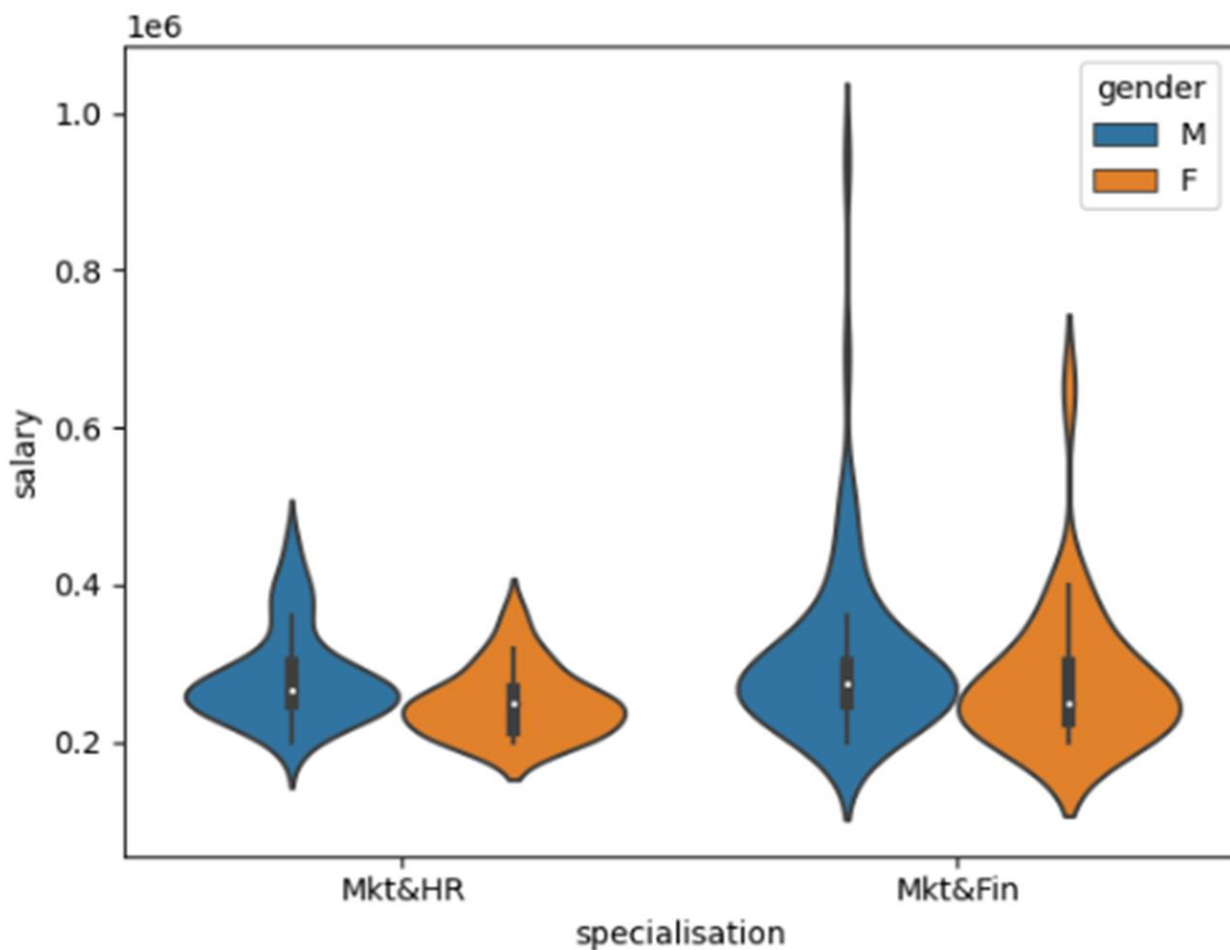
Each plot is explained with relevant examples and insights to help interpret the data effectively.

dataset

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0
...
210	211	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.0	Mkt&Fin	74.49	Placed	400000.0
211	212	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.0	Mkt&Fin	53.62	Placed	275000.0
212	213	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.0	Mkt&Fin	69.72	Placed	295000.0
213	214	F	74.00	Others	66.00	Others	Commerce	58.00	Comm&Mgmt	No	70.0	Mkt&HR	60.23	Placed	204000.0
214	215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.0	Mkt&HR	60.22	Not Placed	NaN

1.Violinplot – Explanation

```
sb.violinplot(x = "specialisation", y = "salary", data = dataset, hue="gender")  
plt.show()
```



Market&Hr - Male&Female:

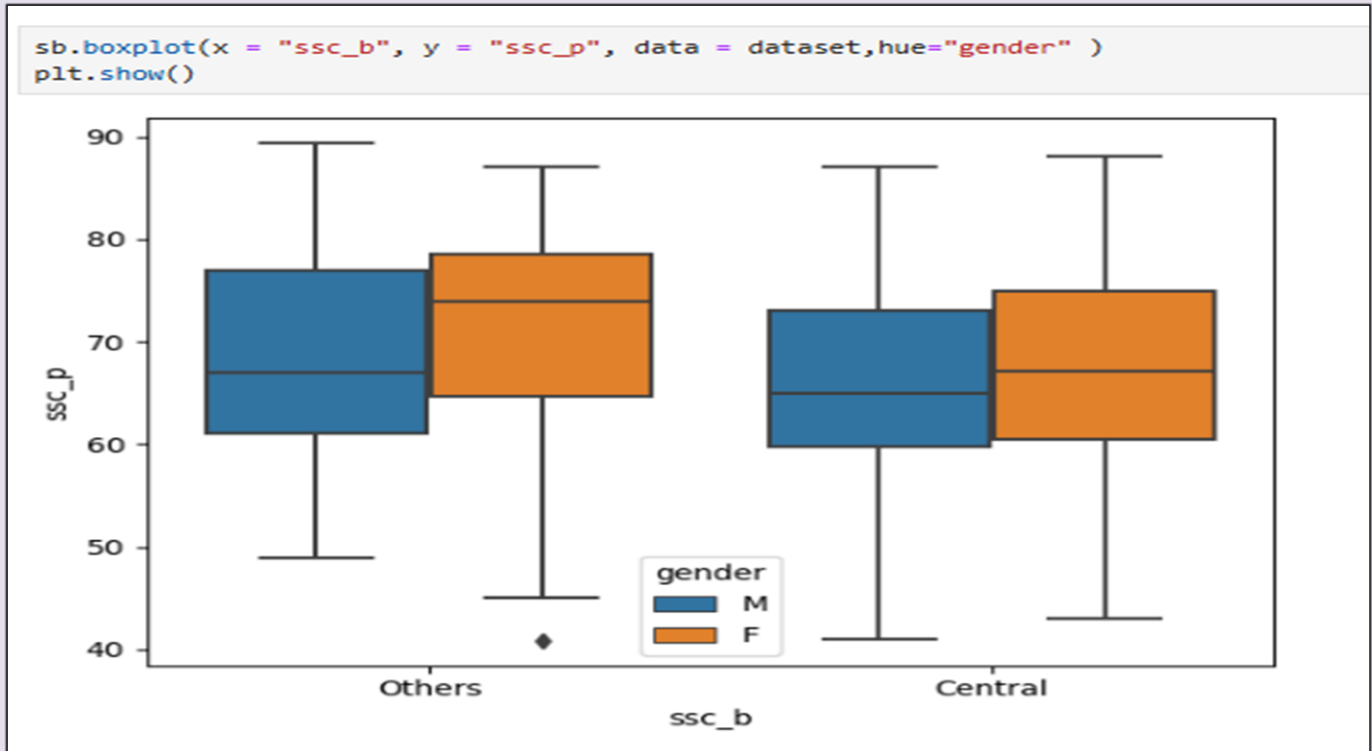
- Male candidates with a minimum salary above 1.25 lakhs are likely to purchase.
- Male candidates with a maximum salary below 5.5 lakhs are likely to purchase.

- ✚ Female candidates with a minimum salary above 1.35 lakhs are likely to purchase.
- ✚ Female candidates with a maximum salary below 4.10 lakhs are likely to purchase.
- ✚ Most male students tend to purchase at a salary level of around 2.50 lakhs.
- ✚ Most female students tend to purchase at a salary level of around 2.30 lakhs.

Market&Finance - Male&Female:

- ✚ Male candidates with a minimum salary above 80 thousand are likely to purchase.
- ✚ Male candidates with a maximum salary below 10 lakhs are likely to purchase.
- ✚ Female candidates with a minimum salary above 80 thousand are likely to purchase.
- ✚ Female candidates with a maximum salary below 7.5 lakhs are likely to purchase.
- ✚ Most male students tend to purchase at a salary level of around 2.80 lakhs.
- ✚ Most female students tend to purchase at a salary level of around 2.30 lakhs.

2.Boxplot - Explanation



Others - Male and Female

- Male candidates scored between 50 and 90 marks. Most candidates scored between 60 and 78 marks, with the average male candidate's score being approximately 67 marks.
- Female candidates scored between 45 and 88 marks. Most candidates scored between 65 and 79 marks, with the average female candidate's score being approximately 75 marks.
- Outlier value presented here.

Central - Male and Female

- # Male candidates scored between 42 and 88 marks. Most candidates scored between 60 and 74 marks, with the average male candidate's score being approximately 64 marks.
- # Female candidates scored between 43 and 89 marks. Most candidates scored between 61 and 73 marks, with the average female candidate's score being approximately 65 marks.

Difference Between Others Male and Central Male

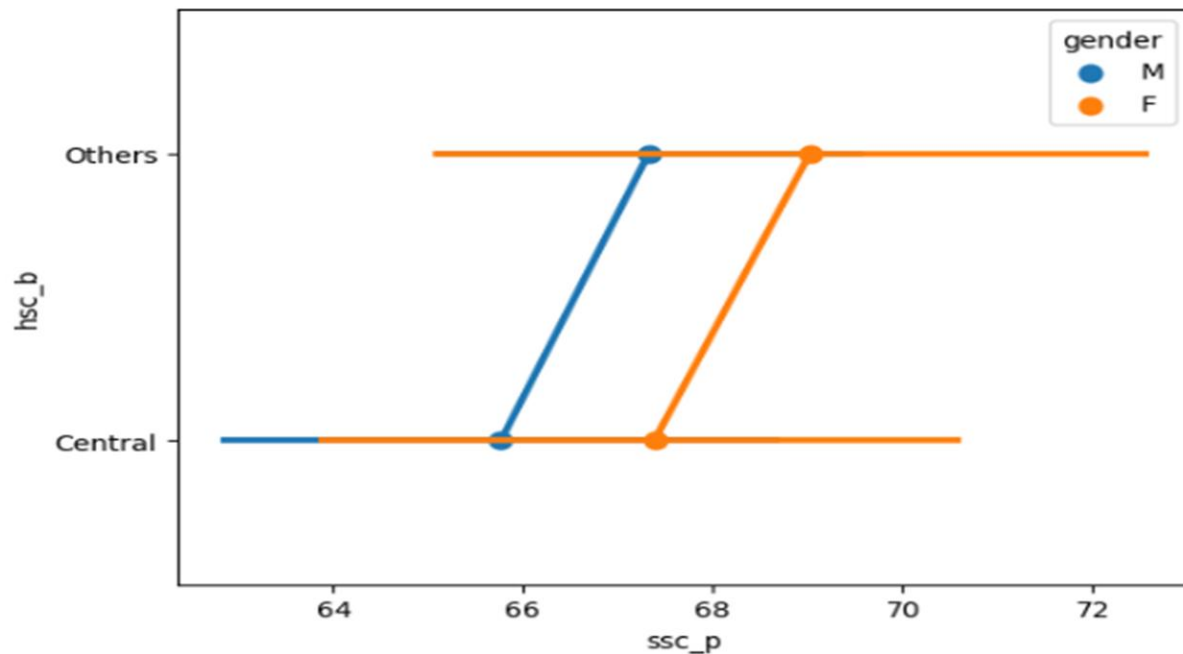
- ❖ Other male candidates have scored better marks than central male candidates. Similarly, other female candidates have scored better marks than central female candidates."

Box plot

- ❖ When interpreting a boxplot, we consider key values such as Q1 (first quartile), Q2 (median), Q3 (third quartile), and the maximum value. Outlier values are not included in the boxplot's main summary.

3.Pointplot - Explanation

```
sb.pointplot(x = "ssc_p",y="hsc_b",data =dataset,hue='gender')  
plt.show()
```



- Male candidates studying in the Others board for 10th standard have scored higher marks than those studying in the Central board.
- Similarly, female candidates studying in the Others board have scored higher marks than those in the Central board.
- Additionally, female candidates have performed better and scored higher marks compared to male candidates overall.

4. Cat Plot - Explanation

```
import seaborn as sb
from matplotlib import pyplot as plt

# Load the dataset
df = sb.load_dataset('exercise')

# Use catplot instead of factorplot
sb.catplot(x="time", y="pulse", hue="kind", kind='violin', col="diet", data=df)
plt.show()
```

C:\Users\anandha rishi\anaconda3\Lib\site-packages\seaborn\categorical.py:641: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.

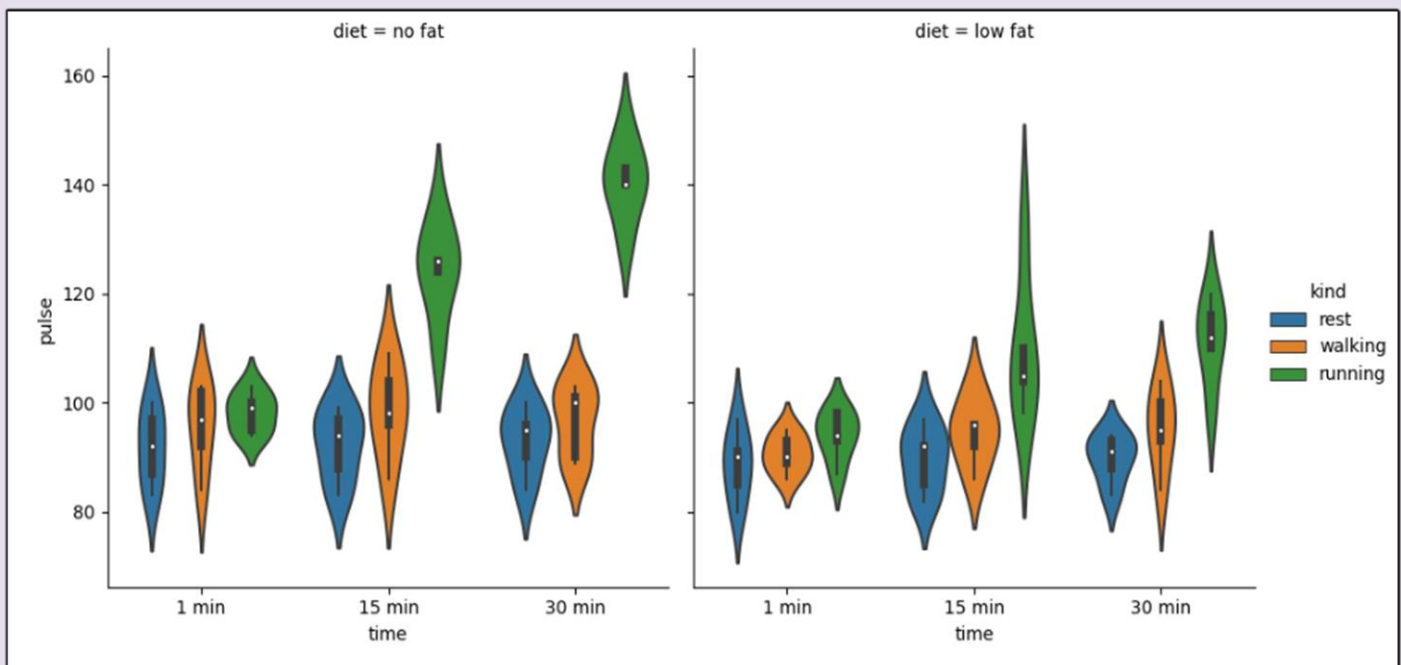
```
grouped_vals = vals.groupby(grouper)
```

C:\Users\anandha rishi\anaconda3\Lib\site-packages\seaborn\categorical.py:641: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.

```
grouped_vals = vals.groupby(grouper)
```

C:\Users\anandha rishi\anaconda3\Lib\site-packages\seaborn\categorical.py:641: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.

```
grouped_vals = vals.groupby(grouper)
```



No Fat Diet:

✚ Rest: For individuals at rest, the pulse rate remains between 70-110 across 1 minute, 15 minutes, and 30 minutes.

✚ Walking: For those walking, the pulse rate stays within 70-120 across 1 minute, 15 minutes, and 30 minutes, with no significant difference.

✚ Running:

1 Minute: Pulse rate is between 90-110.

15 Minutes: Pulse rate ranges between 100-150.

30 Minutes: Pulse rate increases to 120-160.

Low Fat Diet:

✚ Rest: For individuals at rest, the pulse rate is within 70-100 across 1 minute, 15 minutes, and 30 minutes.

✚ Walking: For those walking, the pulse rate stays between 70-110 across all durations, with no major variation.

✚ Running:

1 Minute: Pulse rate ranges from 80-100.

15 Minutes: Pulse rate increases to 80-150.

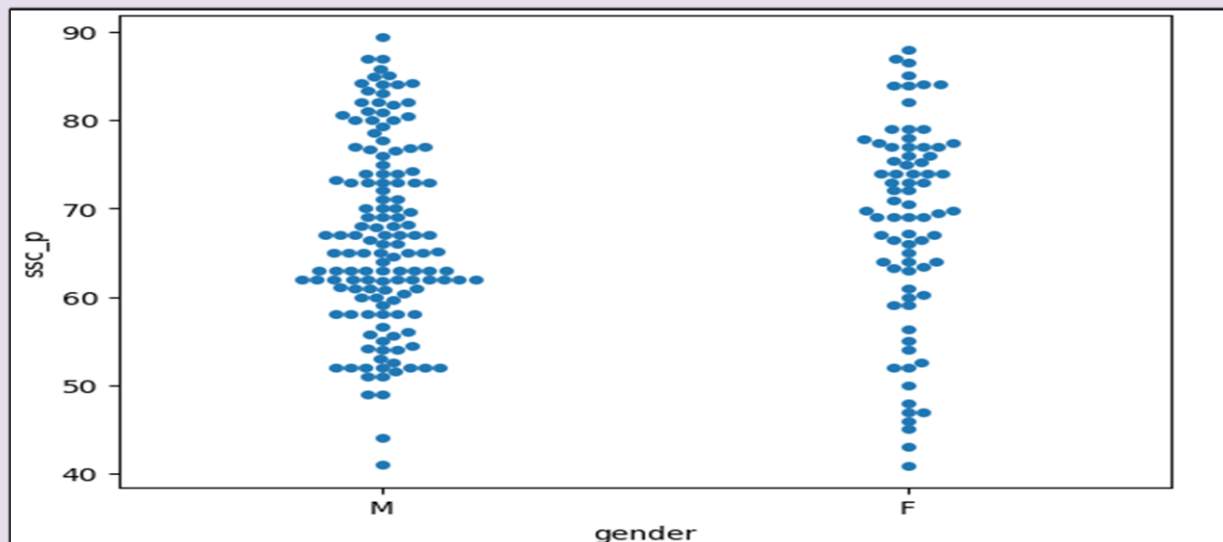
30 Minutes: Pulse rate is within 90-130.

Key Insight:

- ❖ The pulse rate is notably higher only during running.
- ❖ Specifically, individuals on a No Fat Diet exhibit the highest pulse rates during running.

5. Why we use swarm plot

```
sb.swarmplot(x = "gender", y = "ssc_p", data = dataset)  
plt.show()
```



Show Distribution of Data

A swarm plot represents individual data points and their spread for each category. This helps you understand how the data is distributed within each group.

Avoid Overlapping

Unlike a strip plot, the swarm plot adjusts the position of data points to avoid overlap. This makes it easier to see individual data points without them stacking on top of each other.

Combine with Summary Statistics

Swarm plots are often combined with box plots or violin plots. The swarm plot shows the raw data points, while the box or violin plot provides summary statistics like median, quartiles, or density.

Detect Outliers

By showing individual data points, it becomes easier to spot outliers in the dataset.