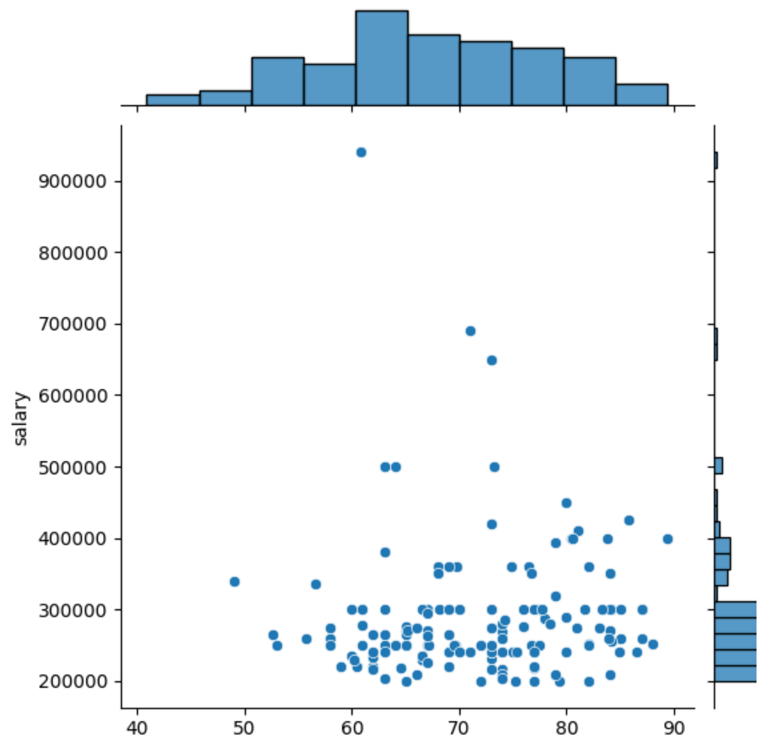
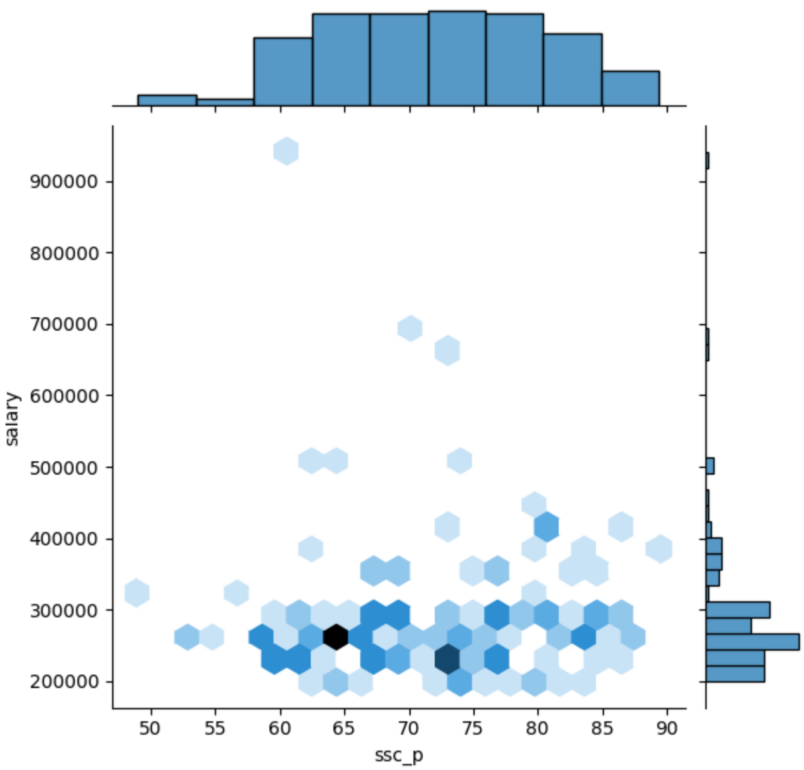
```
#Bivarite Analysis:Distribution Plot
sb.jointplot(x='ssc_p', y='salary',data=dataset)
plt.show()
```




```
#The above plot display a scatterplot with two histograms at the margin of the graphy
#there is positive relationship between this ssc percentage and salary when the marks increase so does the salary.
The scatter plot shows clusters of data points, with the majority concentrated in the marks range of 60-89.
Interestingly, there are two prominent groups
Group 1 (Marks 60-89, Salary 2000000-3500000): The first and more concentrated
group suggests a moderate positive correlation between marks and salary.
Individuals within this range tend to earn salaries ranging from 2000000 to 3500000.
Group 2 (Marks 60-89, Salary 4000000-5000000): The second group, though less dense, exhibits a
moderate positive correlation as well. Individuals in this range earn salaries between 4000000 and 5000000
Notice 1 individual  with mark 60 but a salary exceeding 9000000 indicating that factor beyond academic perfomance
play a significant role in determning earling potential.
The marginal histograms for both marks and salary are left-skewed (negativerly-skewed), indicating that the majority
of data points are concentrated on the right side of their respective distributions.
```

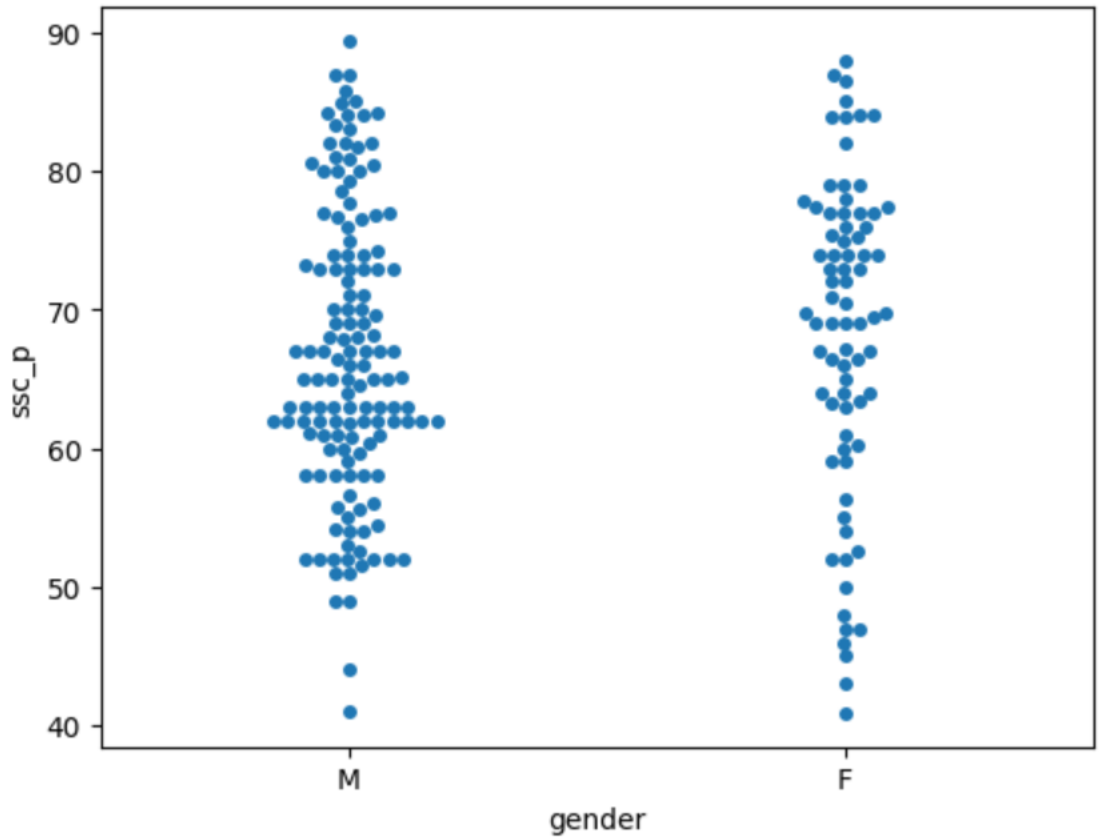
```
#The heatmap offers a comprehensive view of the distribution of data points,
#emphasizing the relationship between academic performance and salary
```

```
sb.jointplot(x='ssc_p', y='salary',data=dataset,kind='hex')
plt.show()
```




```
#Look at the color of each cell to see the strength and direction of the correlation.
#Positive correlations (when one variable increases, the other variable tends to increase) usually represented by warm colors.
#The majority of hexagon bins are densely populated in this region, indicating a moderate positive correlation between marks and salary.
#This region highlights that individuals with marks in the range of 58-89 tend to have salaries ranging from 2000000 to 3000000.
#Darker colors indicate stronger correlations, while lighter colors indicate weaker correlations many individual would have got 65marks n 75marks
```

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



A swarm plot is a type of categorical scatter plot used in data visualization, particularly in Seaborn, a popular Python data visualization library. Swarm plots are often chosen when dealing with categorical data, providing a visual representation of the distribution of individual data points within each category.

The reasons why it is useful:-

Avoid Overlapping Points

Swarm plots are designed to arrange points in a way that avoids overlap, ensuring that each data point is clearly visible. This is particularly important when dealing with discrete or categorical variables where individual data points may share the same or similar values.

Show Distribution

Unlike simple scatter plots where points might overlap, swarm plots provide a clearer indication of the distribution of data within each category. By spreading out the points,

Better Visualization for small – Medium Dataset

Swarm plots work well for relatively small to medium-sized datasets. For larger datasets, the sheer number of points might make the plot too crowded and less informative

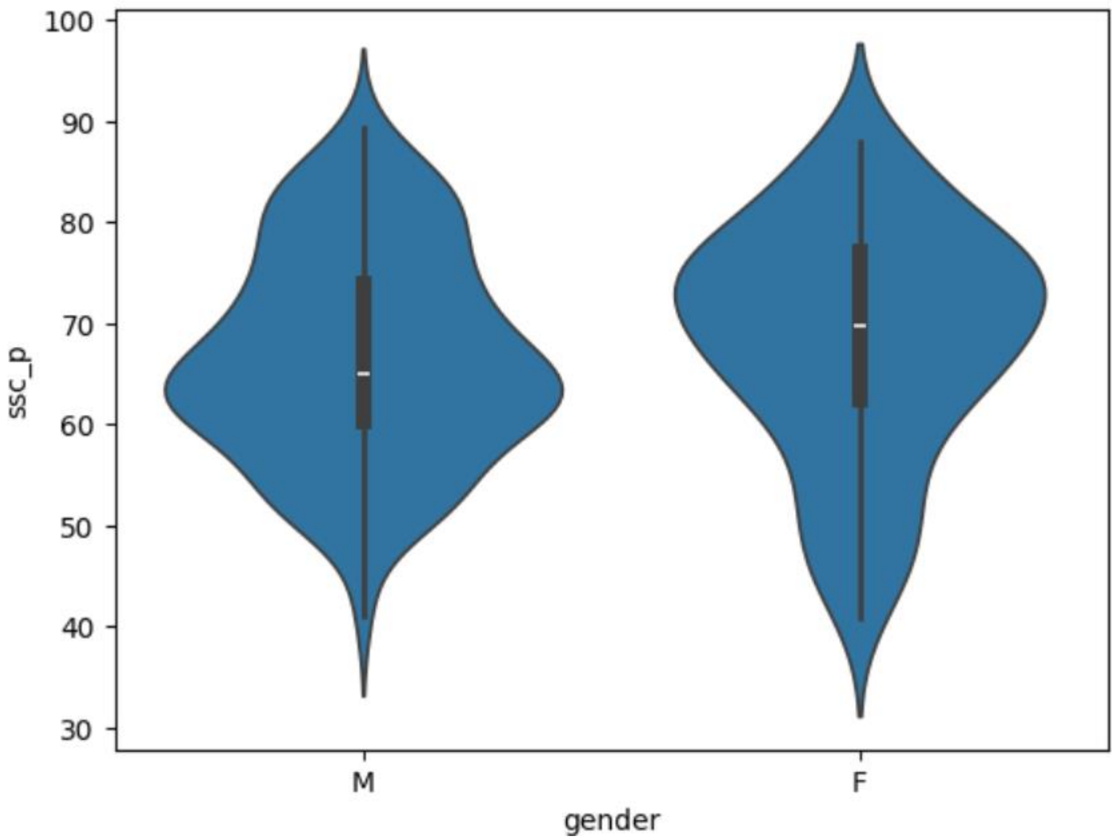
Categorical Comparison

When comparing categories, swarm plots make it easy to see the spread of data points within each category, helping to identify patterns or differences.

Visualize Relationships:

Swarmplots are particularly useful when visualizing relationships between two categorical variables. They help in understanding how the values of one category are distributed concerning the values of another category

```python
sb.violinplot(x='gender', y='ssc_p', data=dataset)
plt.show()
```

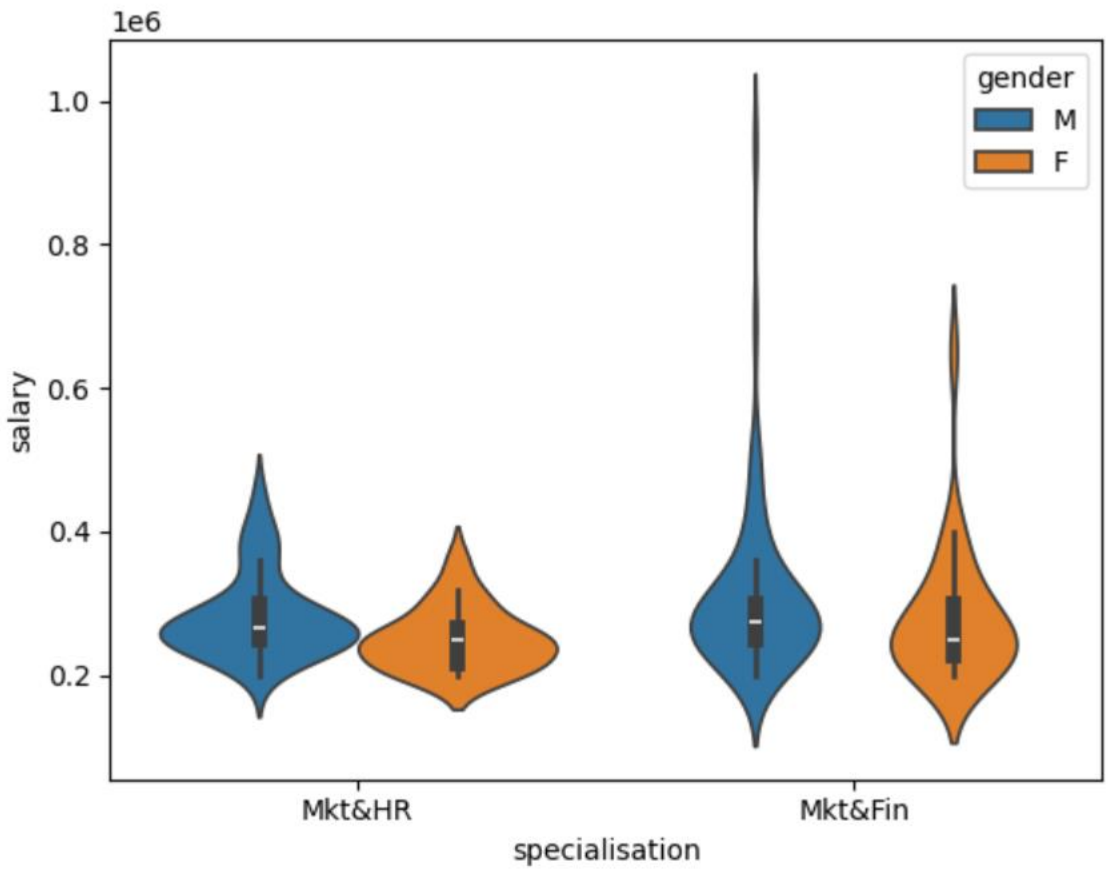


A violin plot is a data visualization that combines aspects of a box plot and a kernel density plot. It is particularly useful for visualizing the distribution and density of data across different categories or groups. The violin plot provides insights into both the summary statistics and the underlying probability density of the data

Box Plot (IQR)S imilar to a traditional box plot, the central part of the violin represents the interquartile range (IQR). The thick line inside the violin indicates the median.

Density Estimation (Kernel Density Plot) The width of the violin plot represents the estimated probability density of the data at different values. It is a smoothed representation of the data distribution. The wider parts of the violin indicate higher density, while the narrower parts indicate lower density

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




```
#A violin plot is a data visualization that combines aspects of a box plot and a kernel density plot
#Visualizing the distribution and density of data across between  Marketing & HR and Marketing & Finance
#groups.'hue' parameter is used to separate male and female in that group.
#In the Marketing & HR department, the male group exhibits a higher salary of 5000000 compare to female with 4000000
#The density of salaries for both genders is relatively similar in the range of 2000000 to 3000000
#As for Marketing & Finance  the male group shows a notable increase in higher salary range surpasing 1000000 whereas
#the female groups salary range stop at 7000000
```