1. Define Problem Statement and perform Exploratory Data Analysis

The main challenge Ola faces is churn rate of its drivers. Ensuring driver loyalty and reducing attrition are crucial to the company's operation.

Analyzing driver data can reveal patterns in driver behavior, performance, and satisfaction. This would help in foreseeing potential churn, allowing proactive measures.

The main aim is to predict potential driver churn using multiple attributes to maintain a consistent driver base and ensure business continuity.

2. Questionnaire :

1. What percentage of drivers have received a quarterly rating of 5 ?

Quartly Rating 5 : No rating as 5 mentioned.
Quartly Rating 4 : 14.27%

```
df[df["Quarterly Rating"]==4]["Driver_ID"].nunique()/df["Driver_ID"].nunique()
```

✓ 0.0s

```
0.14279714405711885
```

2. Comment on the correlation between Age and Quarterly Rating ?

Seems positive Correlation between Age and Quarterly Rating

```
print("\Age <-> Quarterly Rating")
# Calculate Pearson correlation
pearson_corr = df['Age'].corr(df['QR_avg'])
print(f"Pearson Correlation: {pearson_corr}")
# Calculate Spearman correlation
spearman_corr = df['Age'].corr(df['QR_avg'], method='spearman')
print(f"Spearman Correlation: {spearman_corr}")
```

✓ 0.0s

```
\Age <-> Quarterly Rating
Pearson Correlation: 0.19510169050688675
Spearman Correlation: 0.17632230499418575
```

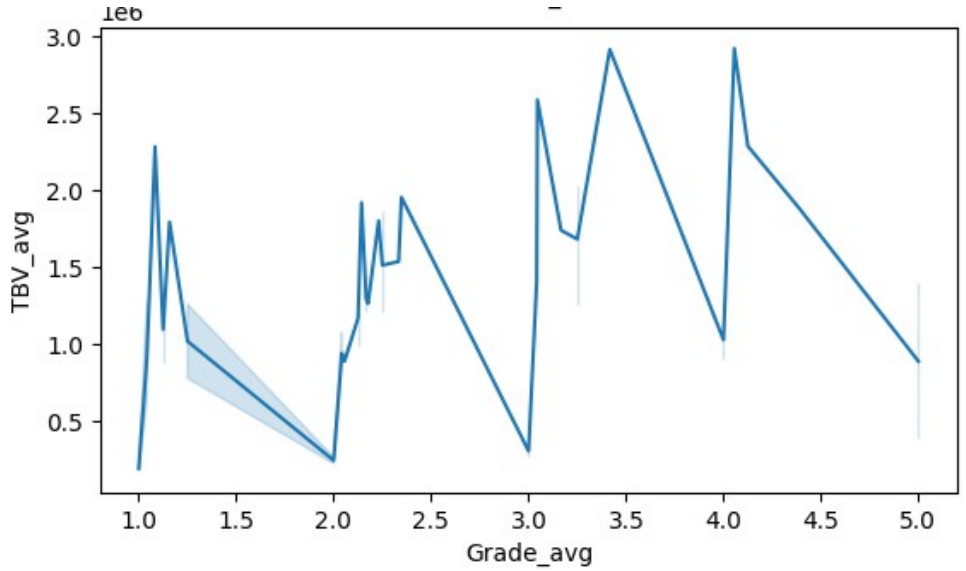
3. Name the city which showed the most improvement in Quarterly Rating over the past year ?

C20 shows the best performance with respect to other Cities.

	City	QR_avg
2	C11	97.775042
8	C17	99.090201
11	C2	103.805939
4	C13	106.959887
28	C9	108.053384
17	C25	111.002885
9	C18	111.980094
10	C19	113.789129
15	C23	116.020759
23	C4	118.006774
16	C24	118.333657
26	C7	120.079781
13	C21	124.682253
0	C1	124.953411
24	C5	127.310358
20	C28	127.866127
5	C14	128.581428
3	C12	128.873701
22	C3	129.639214
25	C6	129.943458
1	C10	135.641087
7	C16	136.358907
14	C22	136.701411
27	C8	140.034518
19	C27	140.872262
6	C15	148.692018
18	C26	159.285706
21	C29	165.080926
12	C20	219.955829

4. Drivers with a Grade of ‘A’ are more likely to have a higher Total Business Value. (T/F)

Grade A/5 : True, likely to have a higher total business values.



5. If a driver's Quarterly Rating drops significantly, how does it impact their Total Business Value in the subsequent period ?

It seems the impact of Quarterly Rating w.r.t Total Business Values
= (1856685 – 82363.69) / 1856685
= 95% drop in Business if rating drops from 4 to 1.

```
Quarterly Rating
1    8.236369e+04
2    5.353634e+05
3    9.358211e+05
4    1.856685e+06
Name: Total Business Value, dtype: float64
```

6. From Ola's perspective, which metric should be the primary focus for driver retention?

1. ROC AUC : It measures the ability of a model to discriminate between positive and negative examples. In the context of driver retention, a high ROC AUC indicates that the model is effective in distinguishing between retained and non-retained drivers. Hence it is an important matrics to consider.

2. Precision : In the context of driver retention, precision would measure how many of the predicted retained drivers are actually retained. It is mainly impactful in understanding whether the Driver continues with us or not.

3. Recall : In the context of driver retention, recall measures the ability of the model to correctly identify retained drivers among all actual retained drivers.
Hence in terms of attrition Recall is the most important metric in comparison to rest.

4. F1 Score : The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall. Hence in terms of overall model sucess it is important.

Conclusion : Recall is the most important metrics, in understanding the Driver’s willingness to continue or not. If we correctly able to understand and catch the Driver’s willingness to leave, we can cater it on time and stop him form leaving the organization.

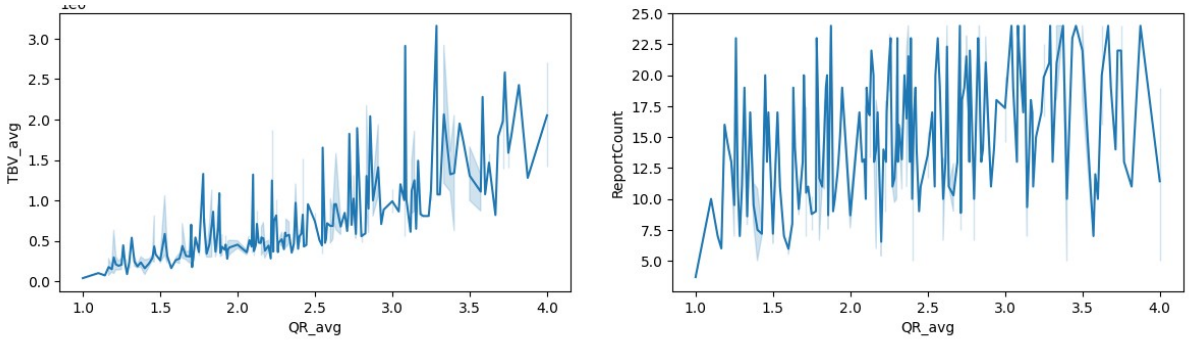
7. How does the gap in precision and recall affect Ola's relationship with its drivers and customers?

Precision: Focuses on more on Driver’s STAYS or not, hence a good precision is helpful in understading the stability of the Business model. With high precision value Ola builds a strong Business foundation and can focus more on Business model. It shows that the decision taken in past were right.

Recall: Focuses more on Driver’s LEAVING or not, a good recall is important for the OLA business model, since it helps in understanding the coming challenges and attrition and can handle them effectly through change in Business model or coming future decisions.

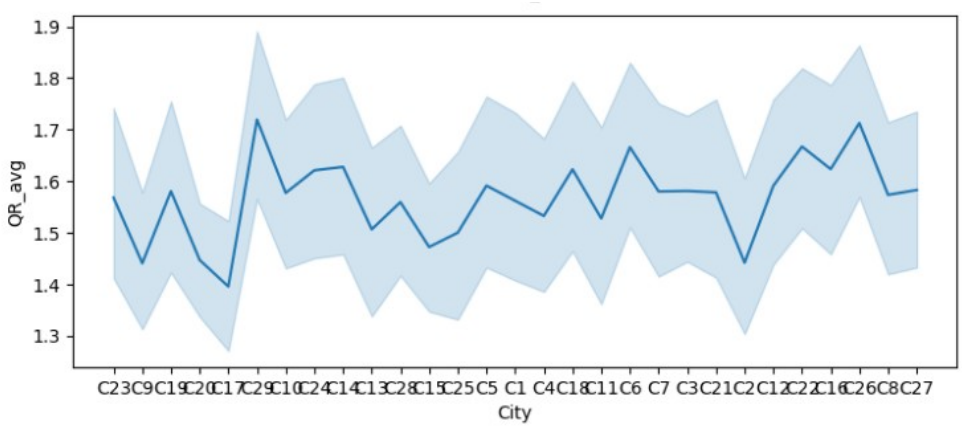
8. Besides the obvious features like "Number of Rides", which lesser-discussed features might have a strong impact on a driver's Quarterly Rating?

I think besides “Number of Rides”, the Total Business Value and Report Counts plays a vital role on Driver’s Quartely Rating.



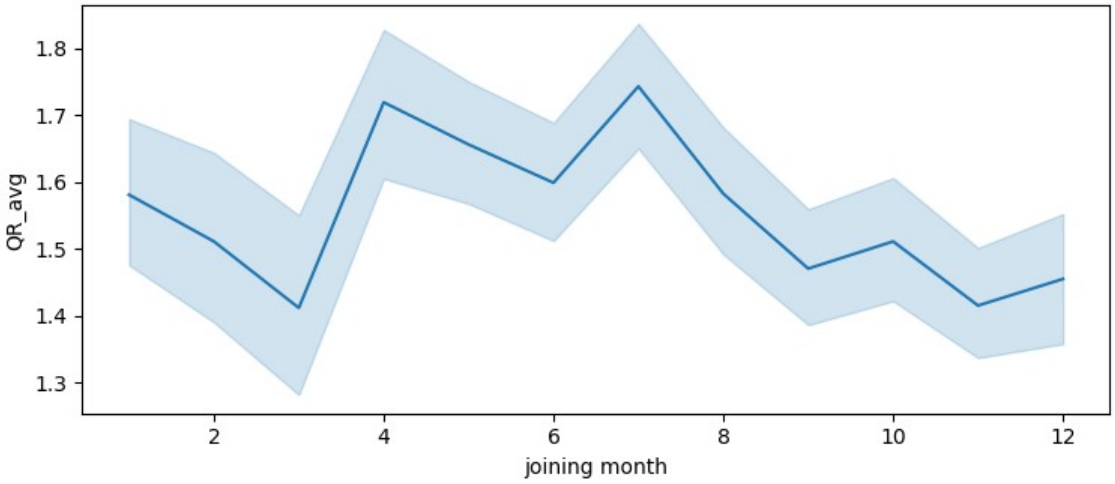
9. Will the driver's performance be affected by the City they operate in? (Yes/No)

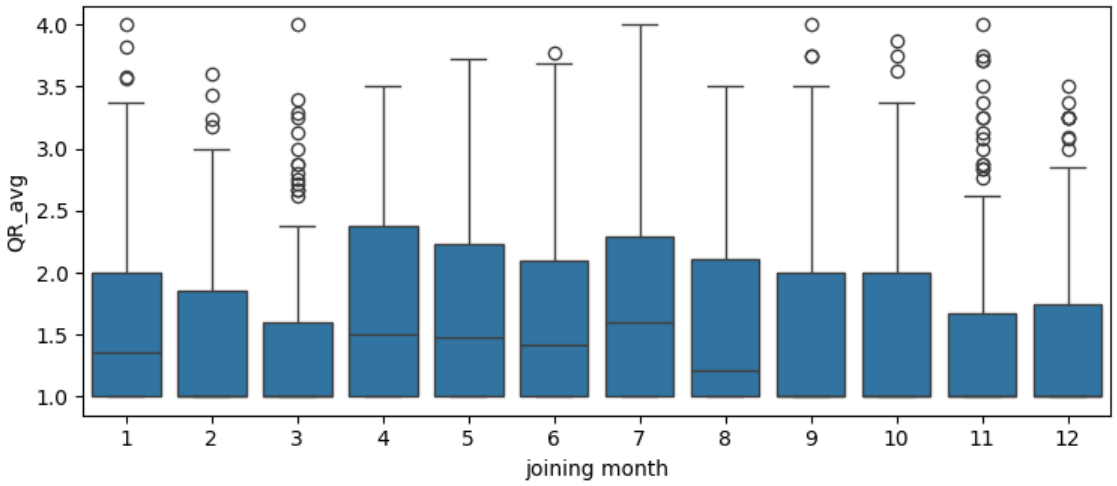
No, it doesn’t seem that City impacts Driver’s performance.



10. Analyze any seasonality in the driver's ratings. Do certain times of the year correspond to higher or lower ratings, and why might that be?

It seems a little positive from April to July, ratings do gets high in these times.
Reasons:
1. In north India it’s Summer time so may be the Driver’s rating goes up as people uses the facilities more.
2. In Summer because of less Traffic on Road the Driver reaches on time and gets better rating as compare to other time.


QR_avg
1.8
1.7
1.6
1.5
1.4
1.3
2
4
6
8
10
12
joining month


QR_avg
4.0
3.5
3.0
2.5
2.0
1.5
1.0
1
2
3
4
5
6
7
8
9
10
11
12
joining month

< END >