FAASOS EDA AND DATA VISUALISATION

- presented by: Anjali K.
- Last updated on: 26th July

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading all my data

In [3]:
```
customers=pd.read_excel("faasos_customers.xls")
driver_orders=pd.read_excel("faasos_driver_order.xls")
drivers=pd.read_excel("faasos_drivers.xls")
ingredients=pd.read_excel("faasos_ingredients.xls")
rolls_recipes=pd.read_excel("faasos_rolls_recipes.xls")
rolls=pd.read_excel("fassos_rolls.xls")
```

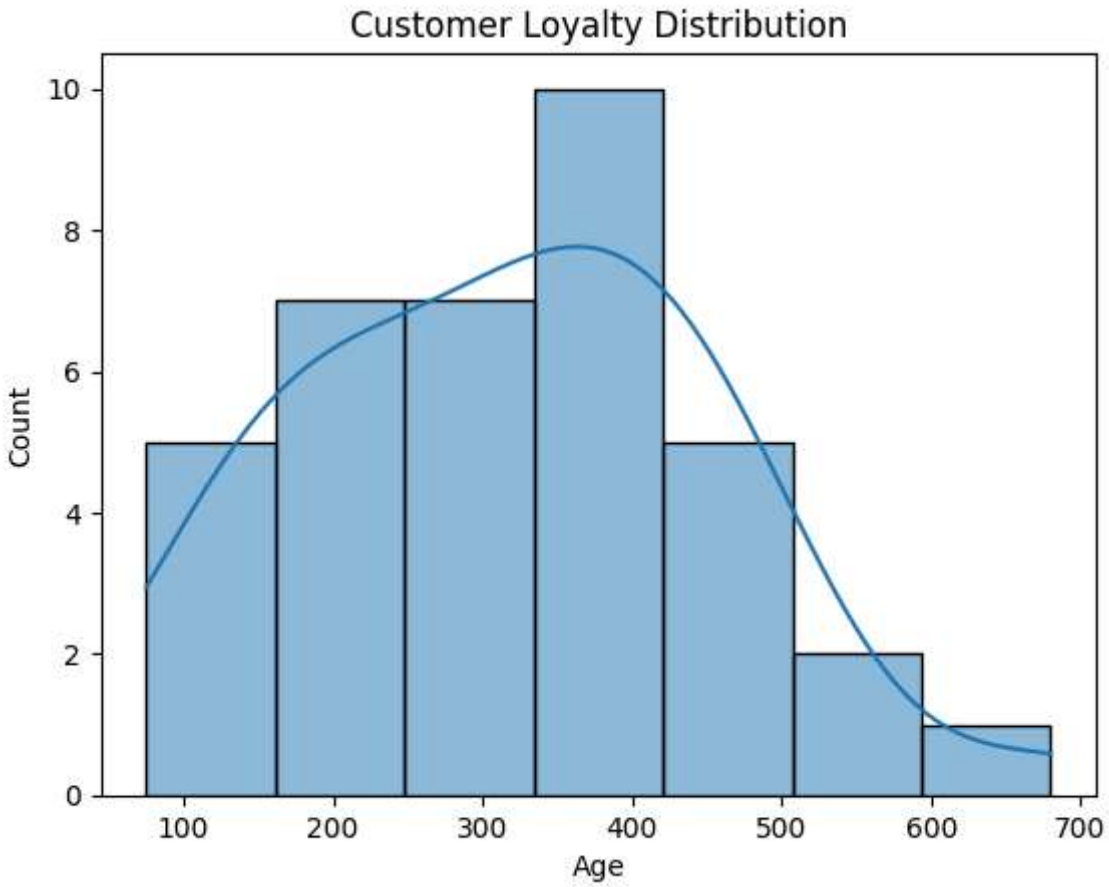
In [4]:
```
customers.head()
```

Out[4]:

	customer_id	customer_name	phone_number	email	address	regist
0	1	Ankit Sharma	-8765432018	ankit.sharma@email.com	123 MG Road, Bangalore	
1	2	Priyanka Singh	-8765432017	priyanka.singh@email.com	456 Park Street, Kolkata	
2	3	Rohit Gupta	-8765432016	rohit.gupta@email.com	789 CP, New Delhi	
3	4	Sneha Patel	-8765432015	sneha.patel@email.com	321 Bandra West, Mumbai	
4	5	Arjun Nair	-8765432014	arjun.nair@email.com	654 Marine Drive, Mumbai	

In [5]:
```
# Plot distribution of customer ages
sns.histplot(customers['loyalty_points'], kde=True)
plt.title("Customer Loyalty Distribution")
plt.xlabel("Age")
```

```
plt.ylabel("Count")
plt.show()
```



This suggests that maximum customers have medium loyalty points

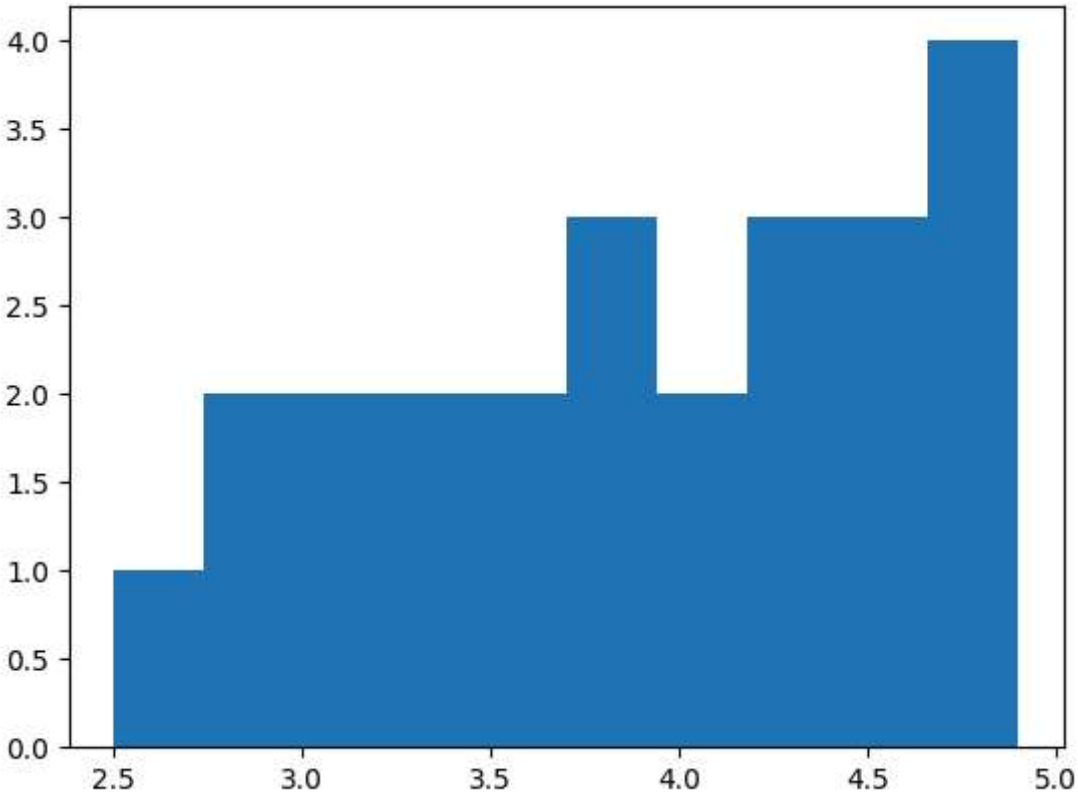
In [6]:
```
drivers.head()
```

Out[6]:

	driver_id	reg_date	driver_name	phone_number	vehicle_type	rating	status
0	1	2021-01-01	Rajesh Kumar	-9876543119	Motorcycle	4.5	Active
1	2	2021-01-03	Amit Singh	-9876543120	Scooter	3.8	Active
2	3	2021-01-08	Priya Sharma	-9876543121	Motorcycle	4.8	Active
3	4	2021-01-15	Mohammed Khan	-9876543122	Bicycle	3.1	Active
4	5	2021-02-01	Sunita Devi	-9876543123	Scooter	4.6	Active

In [7]:
```
plt.hist(drivers.rating)
```

Out[7]:
```
(array([1., 2., 2., 2., 2., 3., 2., 3., 3., 4.]),
 array([2.5 , 2.74, 2.98, 3.22, 3.46, 3.7 , 3.94, 4.18, 4.42, 4.66, 4.9 ]),
 <BarContainer object of 10 artists>)
```



In [8]:
```
driver_orders.head()
```

Out[8]:

	order_id	driver_id	pickup_time	distance	duration	Cancellation	delivery_rating	delive
0	1	1.0	2021-01-01 18:15:34	3.2km	18 minutes	NaN	4.5	
1	2	1.0	2021-01-01 19:10:54	2.8km	15 minutes	NaN	4.2	
2	3	2.0	2021-01-02 23:55:37	5.1km	28 minutes	NaN	4.0	
3	4	2.0	2021-01-02 23:56:15	5.1km	28 minutes	NaN	4.1	
4	5	3.0	2021-01-03 13:35:46	4.7km	22 minutes	NaN	4.8	

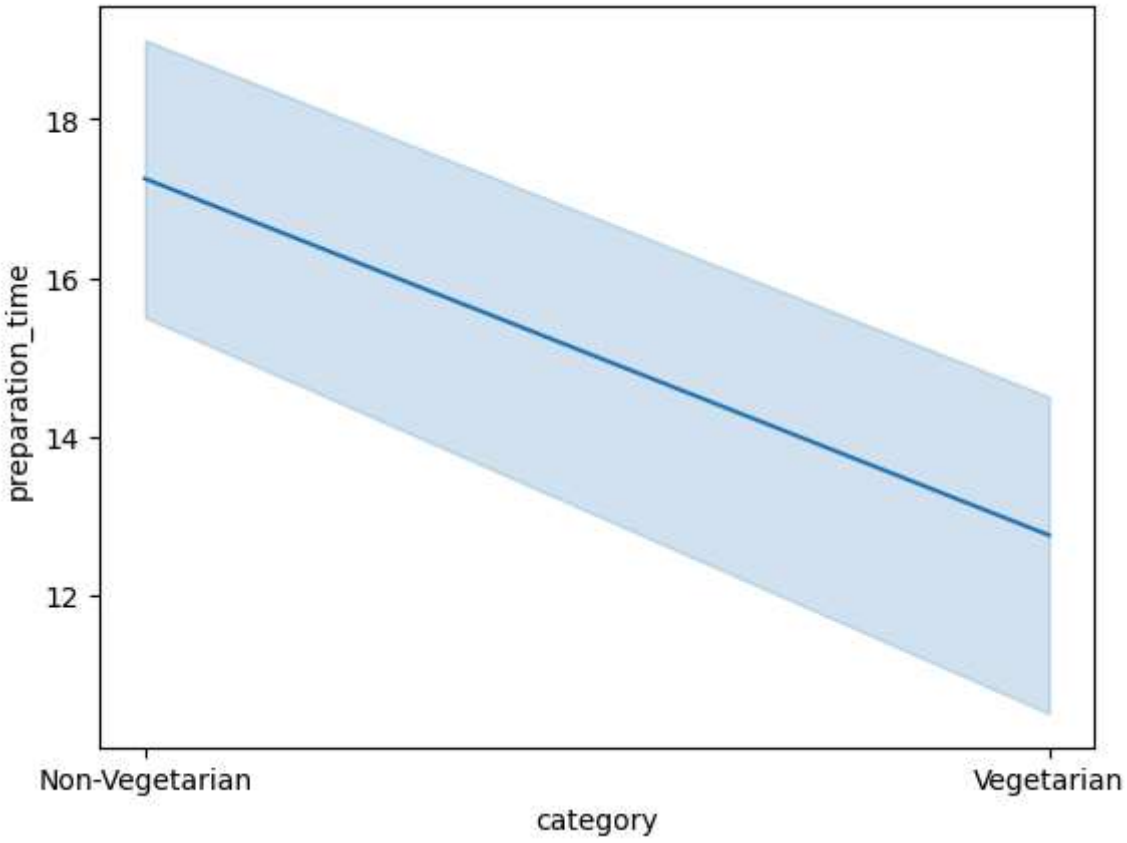
In [9]:
```
rolls.head()
```

Out[9]:

	roll_id	roll_name	base_price	category	preparation_time
0	1	Non Veg Roll	120	Non-Vegetarian	15
1	2	Veg Roll	80	Vegetarian	12
2	3	Paneer Roll	100	Vegetarian	14
3	4	Mutton Roll	180	Non-Vegetarian	20
4	5	Fish Roll	150	Non-Vegetarian	18

In [10]:
```
sns.lineplot(x="category", y="preparation_time", data=rolls)
```

Out[10]:
```
<Axes: xlabel='category', ylabel='preparation_time'>
```



Non Vegetarian rolls are taking more time to prepare

Visulaising the star roll performer

I have taken the data table i need for visualisation from sql and then used it here for further digging deep and visualising the data

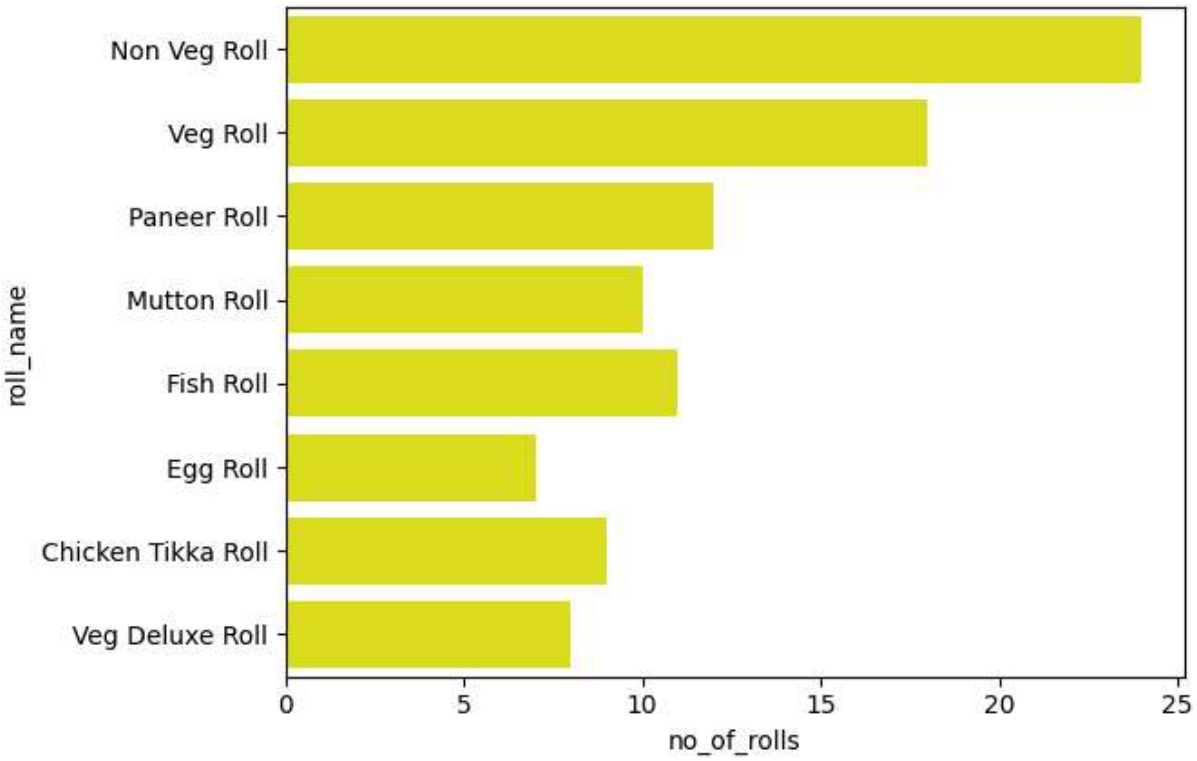
In [13]:
```
roll_performance=pd.read_excel("roll_performance.xls")
roll_performance
```

Out[13]:

	roll_id	roll_name	total_cost	no_of_rolls
0	1	Non Veg Roll	3824	24
1	2	Veg Roll	2131	18
2	3	Paneer Roll	1621	12
3	4	Mutton Roll	1956	10
4	5	Fish Roll	1982	11
5	6	Egg Roll	823	7
6	7	Chicken Tikka Roll	1407	9
7	8	Veg Deluxe Roll	1151	8

In [27]:
```
sns.barplot(x="no_of_rolls",y="roll_name",data=roll_performance,color='yellow')
```

Out[27]:
```
<Axes: xlabel='no_of_rolls', ylabel='roll_name'>
```



- So, non veg roll is the most in demand , followed by veg roll and paneer roll.
- The premium rolls of niche type have less but a steady demand.

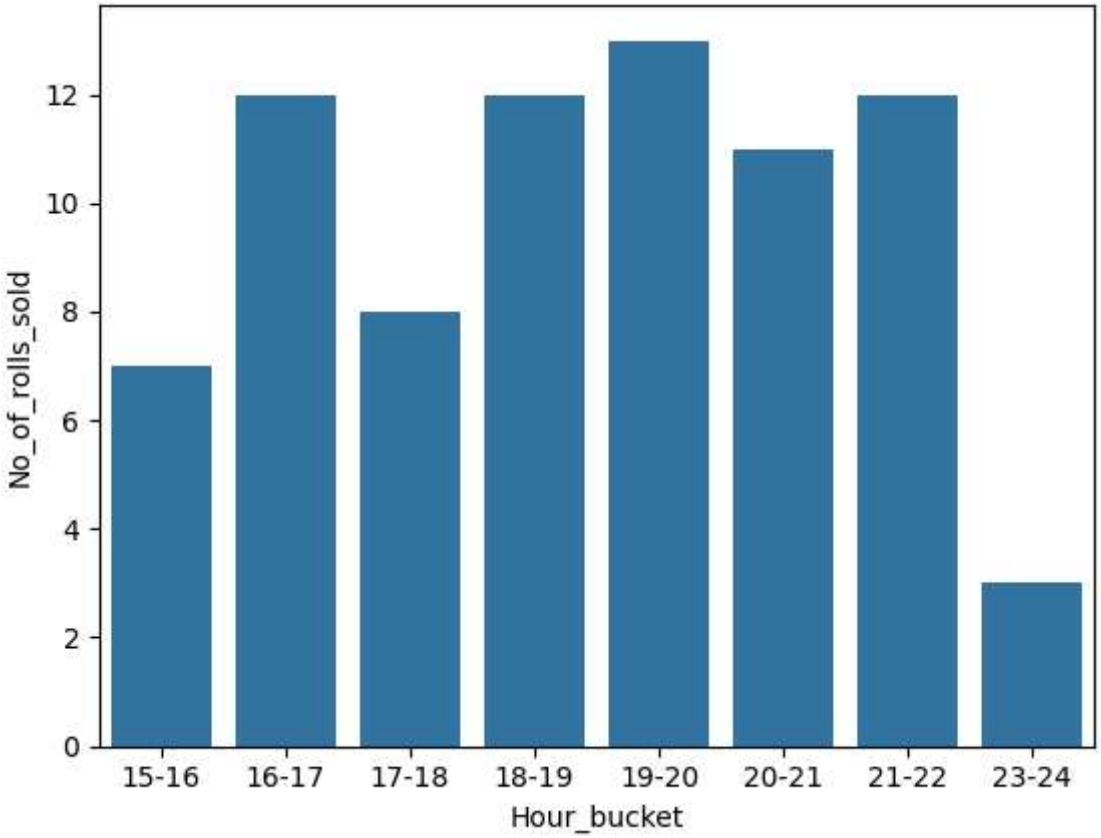
In [65]:
```
hourly_demand=pd.read_excel("hourly_demand.xls")

hourly_demand.rename(columns={"Count ":"No_of_rolls_sold"},inplace=True)
hourly_demand.columns
```

Out[65]:
```
Index(['Hour_bucket', 'No_of_rolls_sold'], dtype='object')
```

In [66]:
```
sns.barplot(x="Hour_bucket",y="No_of_rolls_sold",data=hourly_demand)
```

Out[66]: <Axes: xlabel='Hour_bucket', ylabel='No_of_rolls_sold'>



4pm-5pm, 6pm - 11pm are peak busy hours , more staff can be allocated during these times at headquarters to ensure smooth sales of rolls.

Operational Recommendations:

- Staffing: Increase drivers during 12-14 and 18-21
- Marketing: Target lunch specials (12-14) and late-night offers (21-23)
- Inventory: Prepare for protein-heavy orders (chicken, mutton) in evening