New section

```
import pandas as pd
import seaborn as sns
```

```
sns.get_dataset_names()
```

```
['anagrams',
 'anscombe',
 'attention',
 'brain_networks',
 'car_crashes',
 'diamonds',
 'dots',
 'dowjones',
 'exercise',
 'flights',
 'fmri',
 'geyser',
 'glue',
 'healthexp',
 'iris',
 'mpg',
 'penguins',
 'planets',
 'seaice',
 'taxis',
 'tips',
 'titanic']
```

```
df=sns.load_dataset('flights')
```

```
df
```

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121
...	...	...	...
139	1960	Aug	606
140	1960	Sep	508
141	1960	Oct	461
142	1960	Nov	390
143	1960	Dec	432

144 rows × 3 columns

```
df.head()
```

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121

```
df.tail()
```

	year	month	passengers
139	1960	Aug	606
140	1960	Sep	508
141	1960	Oct	461
142	1960	Nov	390
143	1960	Dec	432

```
df.describe()
```

	year	passengers
count	144.000000	144.000000
mean	1954.500000	280.298611
std	3.464102	119.966317
min	1949.000000	104.000000
25%	1951.750000	180.000000
50%	1954.500000	265.500000
75%	1957.250000	360.500000
max	1960.000000	622.000000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144 entries, 0 to 143
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype   
---  ------      --------------  -----   
 0   year        144 non-null    int64   
 1   month       144 non-null    category
 2   passengers  144 non-null    int64   
dtypes: category(1), int64(2)
memory usage: 2.9 KB
```

```
df.isnull().sum()
```

	0
year	0
month	0
passengers	0

dtype: int64

New Section

```
import pandas as pd
```

```
data=pd.read_csv("Walmart_Sales.csv")
```

```
data
```

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106
...	...	...	...	...	...	...	...	...
6430	45	28-09-2012	713173.95	0	64.88	3.997	192.013558	8.684
6431	45	05-10-2012	733455.07	0	64.89	3.985	192.170412	8.667
6432	45	12-10-2012	734464.36	0	54.47	4.000	192.327265	8.667
6433	45	19-10-2012	718125.53	0	56.47	3.969	192.330854	8.667
6434	45	26-10-2012	760281.43	0	58.85	3.882	192.308899	8.667

6435 rows × 8 columns

```
data.head(10)
```

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	1	05-02-2010	1643690.90	0	42.31	2.572	211.096358	8.106
1	1	12-02-2010	1641957.44	1	38.51	2.548	211.242170	8.106
2	1	19-02-2010	1611968.17	0	39.93	2.514	211.289143	8.106
3	1	26-02-2010	1409727.59	0	46.63	2.561	211.319643	8.106
4	1	05-03-2010	1554806.68	0	46.50	2.625	211.350143	8.106
5	1	12-03-2010	1439541.59	0	57.79	2.667	211.380643	8.106
6	1	19-03-2010	1472515.79	0	54.58	2.720	211.215635	8.106
7	1	26-03-2010	1404429.92	0	51.45	2.732	211.018042	8.106
8	1	02-04-2010	1594968.28	0	62.27	2.719	210.820450	7.808
9	1	09-04-2010	1545418.53	0	65.86	2.770	210.622857	7.808

```
data.describe()
```

	Store	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
count	6435.000000	6.435000e+03	6435.000000	6435.000000	6435.000000	6435.000000	6435.000000
mean	23.000000	1.046965e+06	0.069930	60.663782	3.358607	171.578394	7.999151
std	12.988182	5.643666e+05	0.255049	18.444933	0.459020	39.356712	1.875885
min	1.000000	2.099862e+05	0.000000	-2.060000	2.472000	126.064000	3.879000
25%	12.000000	5.533501e+05	0.000000	47.460000	2.933000	131.735000	6.891000
50%	23.000000	9.607460e+05	0.000000	62.670000	3.445000	182.616521	7.874000
75%	34.000000	1.420159e+06	0.000000	74.940000	3.735000	212.743293	8.622000
max	45.000000	3.818686e+06	1.000000	100.140000	4.468000	227.232807	14.313000

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6435 entries, 0 to 6434
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Store         6435 non-null   int64
 1   Date          6435 non-null   object
 2   Weekly_Sales  6435 non-null   float64
 3   Holiday_Flag  6435 non-null   int64
 4   Temperature   6435 non-null   float64
 5   Fuel_Price    6435 non-null   float64
 6   CPI           6435 non-null   float64
```

```
 7   Unemployment  6435 non-null   float64
dtypes: float64(5), int64(2), object(1)
memory usage: 402.3+ KB
```

```
data.isnull()
```

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
6430	False	False	False	False	False	False	False	False
6431	False	False	False	False	False	False	False	False
6432	False	False	False	False	False	False	False	False
6433	False	False	False	False	False	False	False	False
6434	False	False	False	False	False	False	False	False

6435 rows × 8 columns

```
data.tail(10)
```

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
6425	45	24-08-2012	718232.26	0	72.62	3.834	191.344887	8.684
6426	45	31-08-2012	734297.87	0	75.09	3.867	191.461281	8.684
6427	45	07-09-2012	766512.66	1	75.70	3.911	191.577676	8.684
6428	45	14-09-2012	702238.27	0	67.87	3.948	191.699850	8.684
6429	45	21-09-2012	723086.20	0	65.32	4.038	191.856704	8.684
6430	45	28-09-2012	713173.95	0	64.88	3.997	192.013558	8.684
6431	45	05-10-2012	733455.07	0	64.89	3.985	192.170412	8.667
6432	45	12-10-2012	734464.36	0	54.47	4.000	192.327265	8.667
6433	45	19-10-2012	718125.53	0	56.47	3.969	192.330854	8.667
6434	45	26-10-2012	760281.43	0	58.85	3.882	192.308899	8.667

New Section

```
import pandas as pd
```

```
data=pd.read_csv("insurance.csv")
```

Double-click (or enter) to edit

```
data
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1338 non-null   int64
 1   sex       1338 non-null   object
 2   bmi       1338 non-null   float64
 3   children  1338 non-null   int64
 4   smoker    1338 non-null   object
 5   region    1338 non-null   object
 6   charges   1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

```
data.shape
```

```
(1338, 7)
```

```
data.isnull()
```

	age	sex	bmi	children	smoker	region	charges
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
1333	False	False	False	False	False	False	False
1334	False	False	False	False	False	False	False
1335	False	False	False	False	False	False	False
1336	False	False	False	False	False	False	False
1337	False	False	False	False	False	False	False

1338 rows × 7 columns

```
data.isnull().sum()
```

	0
age	0
sex	0
bmi	0
children	0
smoker	0
region	0
charges	0

dtype: int64

```
data.duplicated()
```

	0
0	False
1	False
2	False
3	False
4	False
...	...
1333	False
1334	False
1335	False
1336	False
1337	False

1338 rows × 1 columns

dtype: bool

```
Start coding or generate with AI.
```

New Section

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

New Section

```
data=pd.read_csv("order_history_kaggle_data.csv")
```

```
data
```

	Restaurant ID	Restaurant name	Subzone	City	Order ID	Order Placed At	Order Status	Delivery	Distance	Items in order	...	Rating	Revi
0	20320607	Swaad	Sector 4	Delhi NCR	6168884918	11:38 PM, September 10 2024	Delivered	Zomato Delivery	3km	1 x Grilled Chicken Jamaican Tender, 1 x Grill...	...	NaN	N
1	20320607	Swaad	Sector 4	Delhi NCR	6170707559	11:34 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Peri Peri Fries, 1 x Fried Chicken Angara ...	...	NaN	N
2	20320607	Swaad	Sector 4	Delhi NCR	6169375019	03:52 PM, September 10 2024	Delivered	Zomato Delivery	<1km	1 x Bone in Peri Peri Grilled Chicken	...	NaN	N
3	20320607	Swaad	Sector 4	Delhi NCR	6151677434	03:45 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Fried Chicken Ghostbuster Tender, 1 x Anga...	...	4.0	N
4	20320607	Swaad	Sector 4	Delhi NCR	6167540897	03:04 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Peri Peri Krispers, 1 x Fried Chicken Anga...	...	NaN	N
...	...	...	...	...	...	...	...	...	...	...	...	...	
21316	21467440	The Chicken Junction	Chittaranjan Park	Delhi NCR	6568490993	03:26 AM, January 30 2025	Delivered	Zomato Delivery	5km	1 x Fried Chicken Desi Peri Peri Bites (Bone)	...	NaN	N
21317	21467440	The Chicken Junction	Chittaranjan Park	Delhi NCR	6564842673	02:44 AM, January 29 2025	Delivered	Zomato Delivery	5km	1 x Fried Chicken 65 Masala Boneless Bites	...	NaN	N
21318	21523055	Masala Junction	Sector 4	Delhi NCR	6553527557	10:05 PM, January 24 2025	Delivered	Zomato Delivery	8km	1 x Korean Fried Chicken Slider	...	5.0	Yum fo Delicio food! smacki food
21319	21523055	Masala Junction	Sector 4	Delhi NCR	6541170157	02:27 PM, January 21 2025	Delivered	Zomato Delivery	9km	1 x Ooh Saucy Fries	...	4.0	The p peri fr we truly tasty a d
21320	21523055	Masala Junction	Sector 4	Delhi NCR	6543973010	02:55 AM, January 21 2025	Delivered	Zomato Delivery	3km	1 x Grilled Chicken Jamaican Slider	...	NaN	N

21321 rows × 29 columns

```
data.head()
```

	Restaurant ID	Restaurant name	Subzone	City	Order ID	Order Placed At	Order Status	Delivery	Distance	Items in order	...	Rating	Review	Cance / Re
0	20320607	Swaad	Sector 4	Delhi NCR	6168884918	11:38 PM, September 10 2024	Delivered	Zomato Delivery	3km	1 x Grilled Chicken Jamaican Tender, 1 x Grill...	...	NaN	NaN	
1	20320607	Swaad	Sector 4	Delhi NCR	6170707559	11:34 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Peri Peri Fries, 1 x Fried Chicken Angara ...	...	NaN	NaN	
2	20320607	Swaad	Sector 4	Delhi NCR	6169375019	03:52 PM, September 10 2024	Delivered	Zomato Delivery	<1km	1 x Bone in Peri Peri Grilled Chicken	...	NaN	NaN	
3	20320607	Swaad	Sector 4	Delhi NCR	6151677434	03:45 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Fried Chicken Ghostbuster Tender, 1 x Anga...	...	4.0	NaN	
4	20320607	Swaad	Sector 4	Delhi NCR	6167540897	03:04 PM, September 10 2024	Delivered	Zomato Delivery	2km	1 x Peri Peri Krispers, 1 x Fried Chicken Anga...	...	NaN	NaN	

5 rows × 29 columns

```
data.tail()
```

	Restaurant ID	Restaurant name	Subzone	City	Order ID	Order Placed At	Order Status	Delivery	Distance	Items in order	...	Rating	Review	C
21316	21467440	The Chicken Junction	Chittaranjan Park	Delhi NCR	6568490993	03:26 AM, January 30 2025	Delivered	Zomato Delivery	5km	1 x Fried Chicken Desi Peri Peri Bites (Bone)	...	NaN	NaN	
21317	21467440	The Chicken Junction	Chittaranjan Park	Delhi NCR	6564842673	02:44 AM, January 29 2025	Delivered	Zomato Delivery	5km	1 x Fried Chicken 65 Masala Boneless Bites	...	NaN	NaN	
21318	21523055	Masala Junction	Sector 4	Delhi NCR	6553527557	10:05 PM, January 24 2025	Delivered	Zomato Delivery	8km	1 x Korean Fried Chicken Slider	...	5.0	Yummy food! Delicious food! Lip smacking food!...	
21319	21523055	Masala Junction	Sector 4	Delhi NCR	6541170157	02:27 PM, January 21 2025	Delivered	Zomato Delivery	9km	1 x Ooh Saucy Fries	...	4.0	The peri peri fries were truly so tasty and de...	
21320	21523055	Masala Junction	Sector 4	Delhi NCR	6543973010	02:55 AM, January 21 2025	Delivered	Zomato Delivery	3km	1 x Grilled Chicken Jamaican Slider	...	NaN	NaN	

5 rows × 29 columns

```
data.describe()
```

	Restaurant ID	Order ID	Bill subtotal	Packaging charges	Restaurant discount (Promo)	Restaurant discount (Flat offs, Freebies & others)	Gold discount	Brand pack discount	Total	
count	2.132100e+04	2.132100e+04	21321.000000	21321.000000	21321.000000	21321.000000	21321.000000	21321.000000	21321.000000	24
mean	2.074413e+07	6.354622e+09	750.076838	32.564592	65.091816	31.795058	0.099128	3.039324	682.616113	
std	2.447193e+05	1.230263e+08	498.759428	22.235898	85.401604	131.487091	3.264261	17.070780	465.313977	
min	2.032061e+07	6.086767e+09	50.000000	0.000000	0.000000	0.000000	0.000000	0.000000	52.500000	
25%	2.063570e+07	6.250751e+09	459.000000	18.450000	0.000000	0.000000	0.000000	0.000000	387.450000	
50%	2.065987e+07	6.357715e+09	629.000000	28.450000	80.000000	0.000000	0.000000	0.000000	597.450000	
75%	2.088265e+07	6.456827e+09	899.000000	39.950000	100.000000	0.000000	0.000000	0.000000	837.900000	
max	2.152306e+07	6.573392e+09	16080.000000	603.000000	4020.000000	7787.000000	280.100000	554.800000	12663.000000	

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21321 entries, 0 to 21320
Data columns (total 29 columns):
 #   Column                                             Non-Null Count  Dtype
---  ------                                             --------------  -----
 0   Restaurant ID                                      21321 non-null  int64
 1   Restaurant name                                    21321 non-null  object
 2   Subzone                                            21321 non-null  object
 3   City                                               21321 non-null  object
 4   Order ID                                           21321 non-null  int64
 5   Order Placed At                                    21321 non-null  object
 6   Order Status                                       21321 non-null  object
 7   Delivery                                           21321 non-null  object
 8   Distance                                           21321 non-null  object
 9   Items in order                                     21321 non-null  object
 10  Instructions                                       720 non-null    object
 11  Discount construct                                 15823 non-null  object
 12  Bill subtotal                                      21321 non-null  float64
 13  Packaging charges                                  21321 non-null  float64
 14  Restaurant discount (Promo)                        21321 non-null  float64
 15  Restaurant discount (Flat offs, Freebies & others) 21321 non-null  float64
 16  Gold discount                                      21321 non-null  float64
 17  Brand pack discount                                21321 non-null  float64
 18  Total                                              21321 non-null  float64
 19  Rating                                             2491 non-null   float64
 20  Review                                             296 non-null    object
 21  Cancellation / Rejection reason                    186 non-null    object
 22  Restaurant compensation (Cancellation)             133 non-null    float64
 23  Restaurant penalty (Rejection)                     3 non-null      float64
 24  KPT duration (minutes)                             21026 non-null  float64
 25  Rider wait time (minutes)                          21153 non-null  float64
 26  Order Ready Marked                                 21321 non-null  object
 27  Customer complaint tag                             469 non-null    object
 28  Customer ID                                        21321 non-null  object
dtypes: float64(12), int64(2), object(15)
memory usage: 4.7+ MB
```

```
data.isnull()
```

	Restaurant ID	Restaurant name	Subzone	City	Order ID	Order Placed At	Order Status	Delivery	Distance	Items in order	...	Rating	Review	Cancellation / Rejection reason	(
0	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
1	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
2	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
3	False	False	False	False	False	False	False	False	False	False	...	False	True	True	
4	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
21316	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
21317	False	False	False	False	False	False	False	False	False	False	...	True	True	True	
21318	False	False	False	False	False	False	False	False	False	False	...	False	False	True	
21319	False	False	False	False	False	False	False	False	False	False	...	False	False	True	
21320	False	False	False	False	False	False	False	False	False	False	...	True	True	True	

21321 rows × 29 columns

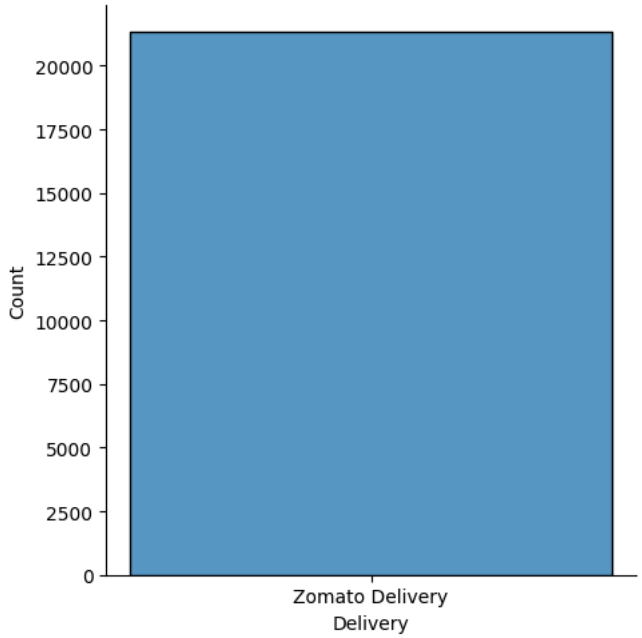
```
data.isnull().sum()
```

	0
Restaurant ID	0
Restaurant name	0
Subzone	0
City	0
Order ID	0
Order Placed At	0
Order Status	0
Delivery	0
Distance	0
Items in order	0
Instructions	20601
Discount construct	5498
Bill subtotal	0
Packaging charges	0
Restaurant discount (Promo)	0
Restaurant discount (Flat offs, Freebies & others)	0
Gold discount	0
Brand pack discount	0
Total	0
Rating	18830
Review	21025
Cancellation / Rejection reason	21135
Restaurant compensation (Cancellation)	21188
Restaurant penalty (Rejection)	21318
KPT duration (minutes)	295
Rider wait time (minutes)	168
Order Ready Marked	0
Customer complaint tag	20852
Customer ID	0

dtype: int64

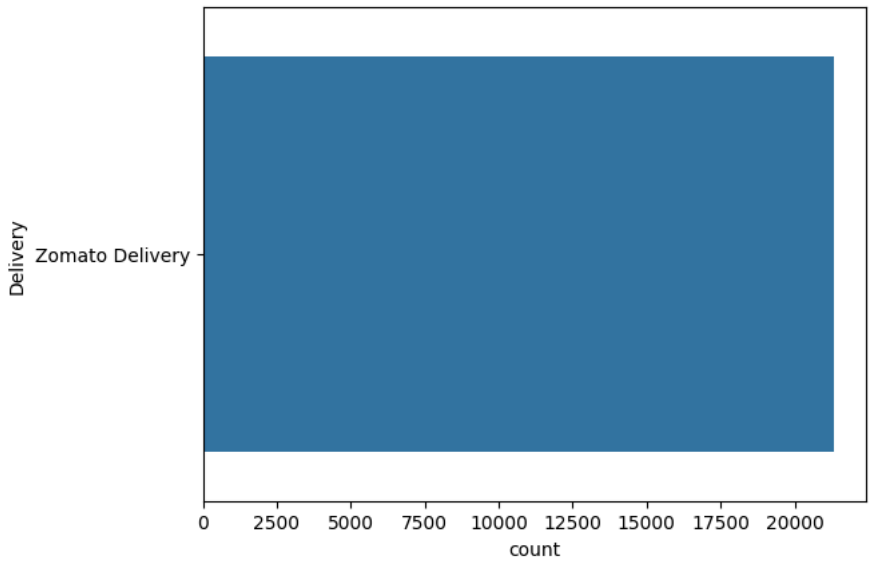
```
sns.displot(data["Delivery"])
```

```
<seaborn.axisgrid.FacetGrid at 0x7d3664d0ac90>
```



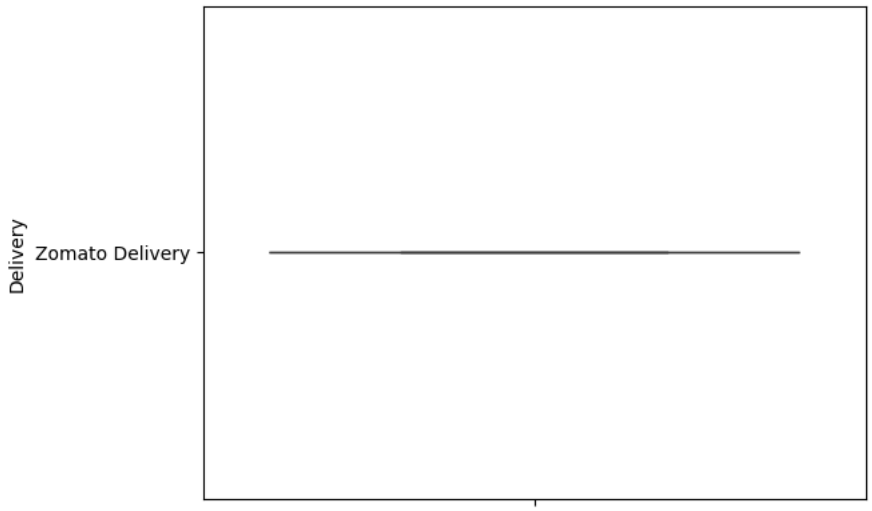

```
sns.countplot(data["Delivery"])
```

```
<Axes: xlabel='count', ylabel='Delivery'>
```



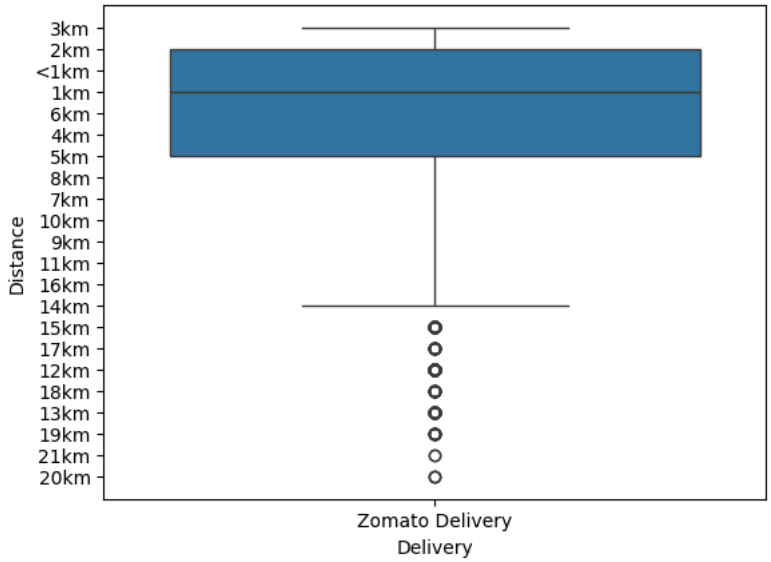

```
sns.boxplot(data["Delivery"])
```

```
<Axes: ylabel='Delivery'>
```




```
sns.boxplot(x="Delivery",y="Distance",data=data)
```

```
<Axes: xlabel='Delivery', ylabel='Distance'>
```




```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data["Delivery"]=le.fit_transform(data["Delivery"])
data["Distance"]=le.fit_transform(data["Distance"])
```

New Section

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

```
data=pd.read_csv("insurance.csv")
```

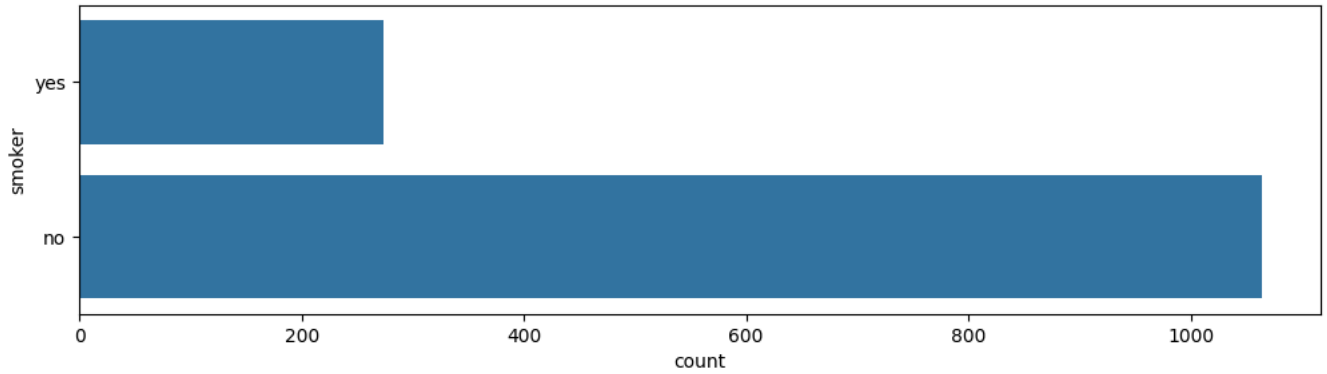
```
data
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

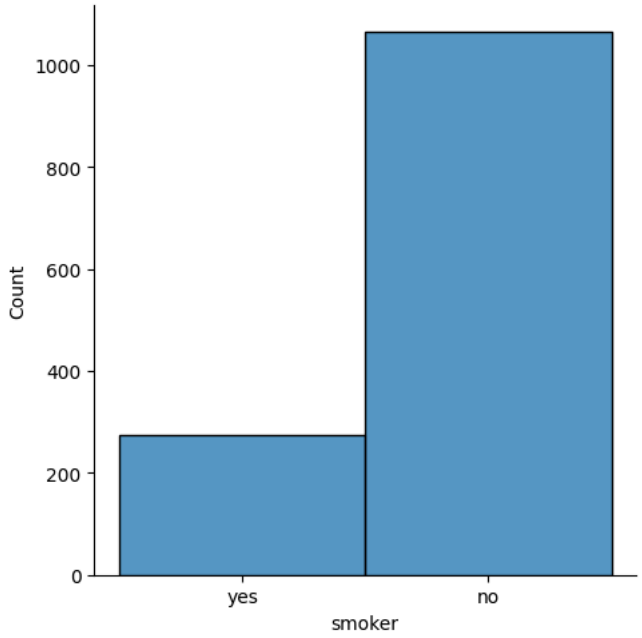
```
plt.figure(figsize=(12,3))
sns.countplot(data["smoker"])
```

```
<Axes: xlabel='count', ylabel='smoker'>
```



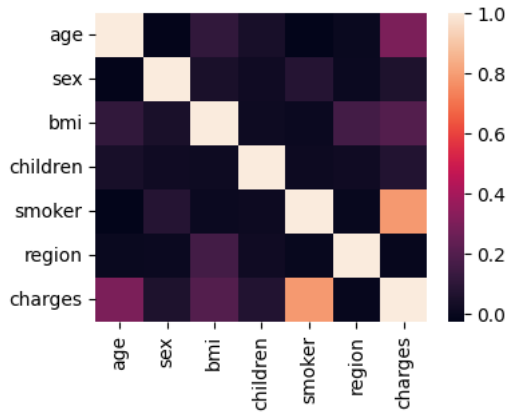

```
plt.figure(figsize=(12,3))
sns.displot(data["smoker"])
```

```
<seaborn.axisgrid.FacetGrid at 0x7d3635d28230>
<Figure size 1200x300 with 0 Axes>
```



```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data["sex"]=le.fit_transform(data["sex"])
data["smoker"]=le.fit_transform(data["smoker"])
data["region"]=le.fit_transform(data["region"])
```

```
plt.figure(figsize=(4,3))
sns.heatmap(data.corr())
plt.show()
```




```
plt.figure(figsize=(8, 4))
sns.histplot(data['age'], kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```