

Task 4

Create visualizations to understand the distribution of variables, identify outliers, and check for correlations between variables.

import the required libraries

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

Import the dataset csv files

```
data1=pd.read_csv('supply_chain_data.csv')
```

data1

	Product type	SKU	Price	Availability	Number of products sold \
0	cosmetics	SKU7	42.958384	59	
426					
1	haircare	SKU81	72.819207	9	
774					
2	skincare	SKU40	80.541424	97	
933					
3	cosmetics	SKU85	76.962994	83	
25					
4	cosmetics	SKU35	84.957787	11	
449					
..
..					
95	cosmetics	SKU92	47.714233	44	
276					
96	haircare	SKU54	31.146243	11	
622					
97	haircare	SKU0	69.808006	55	
802					
98	haircare	SKU30	8.022859	10	
327					
99	skincare	SKU1	14.843523	95	
736					

	Revenue generated	Customer demographics	Stock levels	Lead times
0	8496.103813	Female	93	17
1	4384.413400	Unknown	48	6
2	5724.959350	Female	90	20

3	8684.613059		Female	15	18
4	6541.329345		Female	42	27
..
95	2100.129755		Male	90	25
96	6088.021480		Non-binary	33	22
97	8661.996792		Non-binary	58	7
98	2766.342367		Male	60	26
99	7460.900065		Female	53	30
	Order quantities	...	Location	Lead time	Production volumes \
0	11	...	Bangalore	22	564
1	8	...	Delhi	28	698
2	39	...	Kolkata	18	793
3	66	...	Chennai	4	211
4	85	...	Delhi	3	367
..
95	10	...	Mumbai	4	671
96	61	...	Kolkata	26	497
97	96	...	Mumbai	29	215
98	35	...	Kolkata	27	806
99	37	...	Mumbai	23	517
	Manufacturing lead time		Manufacturing costs	Inspection results	\
0	1		99.466109	Fail	
1	1		19.789593	Pending	
2	1		88.179407	Pending	
3	2		69.929346	Fail	
4	2		58.004787	Pass	
..	
95	29		62.612690	Pass	
96	29		30.186023	Pass	
97	29		46.279879	Pending	
98	30		51.634893	Pending	
99	30		33.616769	Pending	
	Defect rates	Transportation modes	Routes	Costs	
0	0.398177	Road	Route C	802.056312	
1	2.547547	Rail	Route B	276.778336	
2	4.213269	Road	Route A	529.808724	
3	1.374429	Road	Route B	842.686830	
4	0.541154	Sea	Route C	553.420471	

95	0.333432	Rail	Route B	230.092782
96	2.478772	Road	Route B	814.069997
97	0.226410	Road	Route B	187.752075
98	0.965395	Road	Route C	880.080988
99	4.854068	Road	Route B	503.065579

[100 rows x 24 columns]

Understanding the data distribution

data1.head()

	Product type	SKU	Price	Availability	Number of products sold \
0	cosmetics	SKU7	42.958384		59
1	hairecare	SKU81	72.819207		9
2	skincare	SKU40	80.541424		97
3	cosmetics	SKU85	76.962994		83
4	cosmetics	SKU35	84.957787		11

	Revenue generated	Customer demographics	Stock levels	Lead times \
0	8496.103813	Female	93	17
1	4384.413400	Unknown	48	6
2	5724.959350	Female	90	20
3	8684.613059	Female	15	18
4	6541.329345	Female	42	27

	Order quantities	...	Location	Lead time	Production volumes \
0	11	...	Bangalore	22	564
1	8	...	Delhi	28	698
2	39	...	Kolkata	18	793
3	66	...	Chennai	4	211
4	85	...	Delhi	3	367

	Manufacturing lead time	Manufacturing costs	Inspection results \
0	1	99.466109	Fail
1	1	19.789593	Pending
2	1	88.179407	Pending

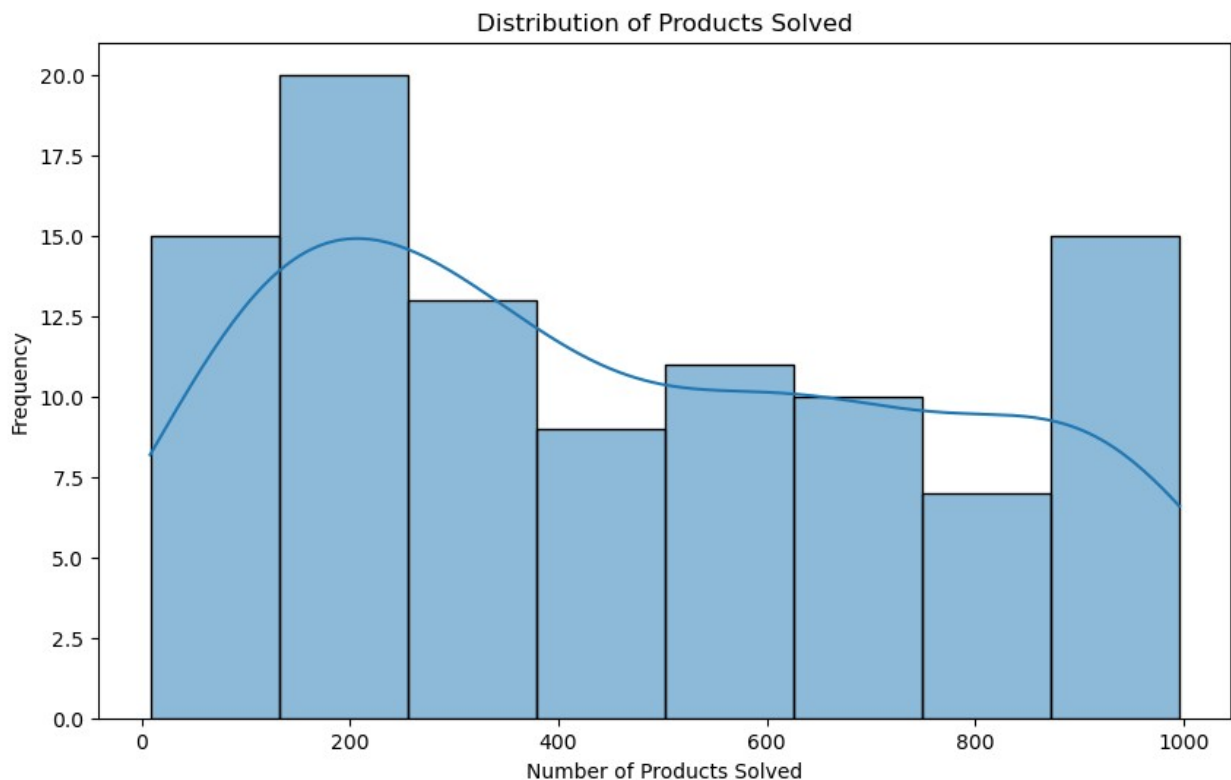
3	2	69.929346	Fail
4	2	58.004787	Pass

	Defect rates	Transportation modes	Routes	Costs
0	0.398177	Road	Route C	802.056312
1	2.547547	Rail	Route B	276.778336
2	4.213269	Road	Route A	529.808724
3	1.374429	Road	Route B	842.686830
4	0.541154	Sea	Route C	553.420471

[5 rows x 24 columns]

Histogram

```
plt.figure(figsize=(10, 6))
sns.histplot(data1['Number of products sold'], kde=True)
plt.title('Distribution of Products Sold')
plt.xlabel('Number of Products Sold')
plt.ylabel('Frequency')
plt.show()
```



```
plt.figure(figsize=(10, 6))
sns.boxplot(x=data1['Revenue generated'])
plt.title('Distribution of Revenue Generated')
```

```
plt.xlabel('Revenue Generated')  
plt.show()
```

