

SupermarketSales

September 15, 2025

```
[36]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\user\Downloads\supermarket_sales - Sheet1.csv")
df.head()
```

```
[36]:
```

	Invoice ID	Branch	City	Customer type	Gender	\
0	750-67-8428	A	Yangon	Member	Female	
1	226-31-3081	C	Naypyitaw	Normal	Female	
2	631-41-3108	A	Yangon	Normal	Male	
3	123-19-1176	A	Yangon	Member	Male	
4	373-73-7910	A	Yangon	Normal	Male	

	Product line	Unit price	Quantity	Tax 5%	Total	Date	\
0	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	
1	Electronic accessories	15.28	5	3.8200	80.2200	3/8/2019	
2	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	
3	Health and beauty	58.22	8	23.2880	489.0480	1/27/2019	
4	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	

	Time	Payment	cogs	gross margin percentage	gross income	Rating
0	13:08	Ewallet	522.83	4.761905	26.1415	9.1
1	10:29	Cash	76.40	4.761905	3.8200	9.6
2	13:23	Credit card	324.31	4.761905	16.2155	7.4
3	20:33	Ewallet	465.76	4.761905	23.2880	8.4
4	10:37	Ewallet	604.17	4.761905	30.2085	5.3

Drop Some Columns

```
[37]: df = df.drop(columns=["Invoice ID", "gross margin percentage"])
```

```
[58]: df.duplicated()
```

```
[58]: 0    False
1    False
2    False
3    False
```

```

4      False
...
995    False
996    False
997    False
998    False
999    False
Length: 1000, dtype: bool

```

Change Date and Time

```
[39]: df['Date'] = pd.to_datetime(df['Date'])
```

```
[59]: df.head()
```

```
[59]:
```

	Branch	City	Customer type	Gender	Product line	Unit price \
0	A	Yangon	Member	Female	Health and beauty	74.69
1	C	Naypyitaw	Normal	Female	Electronic accessories	15.28
2	A	Yangon	Normal	Male	Home and lifestyle	46.33
3	A	Yangon	Member	Male	Health and beauty	58.22
4	A	Yangon	Normal	Male	Sports and travel	86.31

	Quantity	Tax 5%	Total	Date	...	cogs	gross income	Rating \
0	7	26.1415	548.9715	2019-01-05	...	522.83	26.1415	9.1
1	5	3.8200	80.2200	2019-03-08	...	76.40	3.8200	9.6
2	7	16.2155	340.5255	2019-03-03	...	324.31	16.2155	7.4
3	8	23.2880	489.0480	2019-01-27	...	465.76	23.2880	8.4
4	7	30.2085	634.3785	2019-02-08	...	604.17	30.2085	5.3

	Year	Month	Day	DayOfWeek	weekend	Hour	TimeOfDay
0	2019	1	5	Saturday	False	13	Afternoon
1	2019	3	8	Friday	False	10	Morning
2	2019	3	3	Sunday	False	13	Afternoon
3	2019	1	27	Sunday	False	20	Evening
4	2019	2	8	Friday	False	10	Morning

[5 rows x 22 columns]

ADD ANOTHER COLUMNS

Ibi bizadufasha kubona:

Umwaka & ukwezi → trends za sales.

Umunsi wa week → busiest day.

Weekend vs Weekday → kumenya uko sales zihinduka.

```
[41]: df['Year'] =df['Date'].dt.year
df['Month'] = df['Date'].dt.month
```

```
df['Day']=df['Date'].dt.day
df['DayOfWeek']=df['Date'].dt.day_name()
df['weekend']=df['DayOfWeek'].isin(['saturday','sunday'])
```

```
[42]: df.head()
```

```
[42]:
```

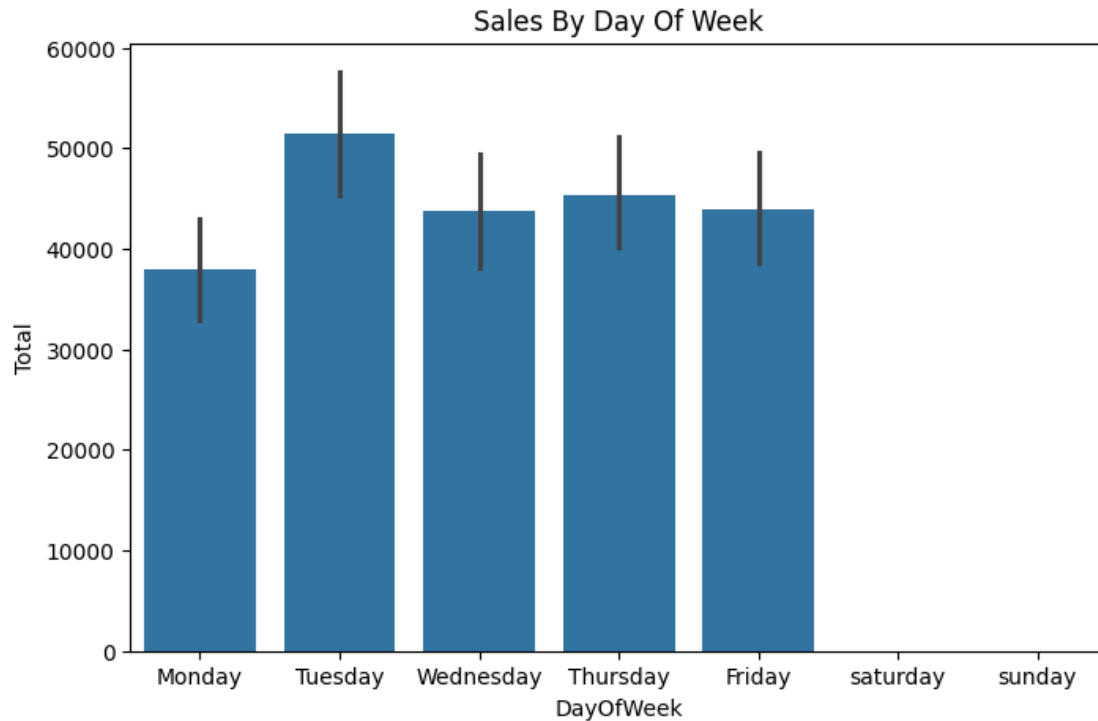
	Branch	City	Customer type	Gender	Product line	Unit price	\
0	A	Yangon	Member	Female	Health and beauty	74.69	
1	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	
2	A	Yangon	Normal	Male	Home and lifestyle	46.33	
3	A	Yangon	Member	Male	Health and beauty	58.22	
4	A	Yangon	Normal	Male	Sports and travel	86.31	

	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	\
0	7	26.1415	548.9715	2019-01-05	13:08	Ewallet	522.83	
1	5	3.8200	80.2200	2019-03-08	10:29	Cash	76.40	
2	7	16.2155	340.5255	2019-03-03	13:23	Credit card	324.31	
3	8	23.2880	489.0480	2019-01-27	20:33	Ewallet	465.76	
4	7	30.2085	634.3785	2019-02-08	10:37	Ewallet	604.17	

	gross income	Rating	Year	Month	Day	DayOfWeek	weekend
0	26.1415	9.1	2019	1	5	Saturday	False
1	3.8200	9.6	2019	3	8	Friday	False
2	16.2155	7.4	2019	3	3	Sunday	False
3	23.2880	8.4	2019	1	27	Sunday	False
4	30.2085	5.3	2019	2	8	Friday	False

Visualization: Sales per Day of Week

```
[44]: plt.figure(figsize=(8,5))
sns.barplot(x='DayOfWeek', y='Total', data=df,
            estimator='sum', order=['Monday','Tuesday','Wednesday','Thursday','Friday','saturday','sunday'])
plt.title("Sales By Day Of Week")
plt.show()
```



#ADD NEW COLUMN FROM TIME

```
[46]: # change Time into dateTime
df['Time'] =pd.to_datetime(df['Time'])

#ADD column HOUR
df['Hour']=df['Time'].dt.hour
```

```
[47]: #ADD column of Time of Day

def get_time_of_day(hour):
    if 5 <= hour < 12:
        return 'Morning'
    elif 12 <= hour < 17:
        return 'Afternoon'
    elif 17 <= hour < 21:
        return 'Evening'
    else:
        return 'Night'

df['TimeOfDay'] = df['Hour'].apply(get_time_of_day)
```

```
[49]: #Ibi biduha busiest hours na busiest time of day
df.head()
```

```
[49]:
```

	Branch	City	Customer type	Gender	Product line	Unit price \
0	A	Yangon	Member	Female	Health and beauty	74.69
1	C	Naypyitaw	Normal	Female	Electronic accessories	15.28
2	A	Yangon	Normal	Male	Home and lifestyle	46.33
3	A	Yangon	Member	Male	Health and beauty	58.22
4	A	Yangon	Normal	Male	Sports and travel	86.31

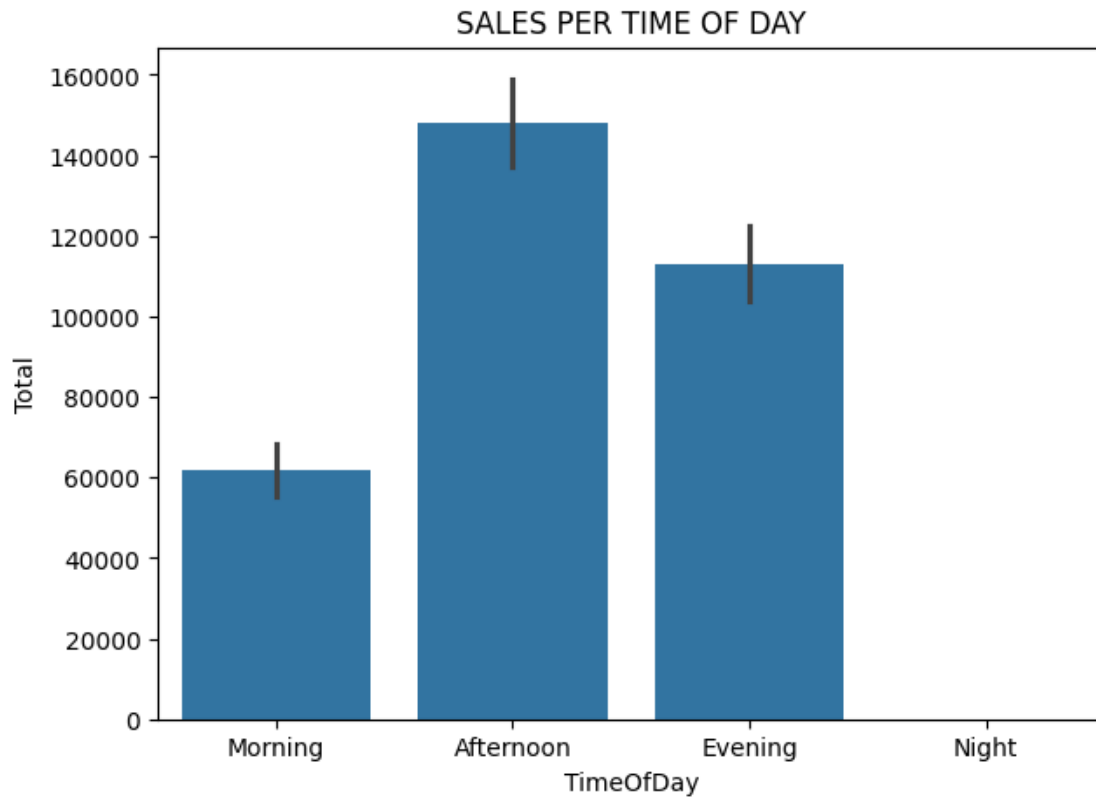
	Quantity	Tax 5%	Total	Date	...	cogs	gross income	Rating \
0	7	26.1415	548.9715	2019-01-05	...	522.83	26.1415	9.1
1	5	3.8200	80.2200	2019-03-08	...	76.40	3.8200	9.6
2	7	16.2155	340.5255	2019-03-03	...	324.31	16.2155	7.4
3	8	23.2880	489.0480	2019-01-27	...	465.76	23.2880	8.4
4	7	30.2085	634.3785	2019-02-08	...	604.17	30.2085	5.3

	Year	Month	Day	DayOfWeek	weekend	Hour	TimeOfDay
0	2019	1	5	Saturday	False	13	Afternoon
1	2019	3	8	Friday	False	10	Morning
2	2019	3	3	Sunday	False	13	Afternoon
3	2019	1	27	Sunday	False	20	Evening
4	2019	2	8	Friday	False	10	Morning

[5 rows x 22 columns]

1 Visualization: Sales per Time of Day

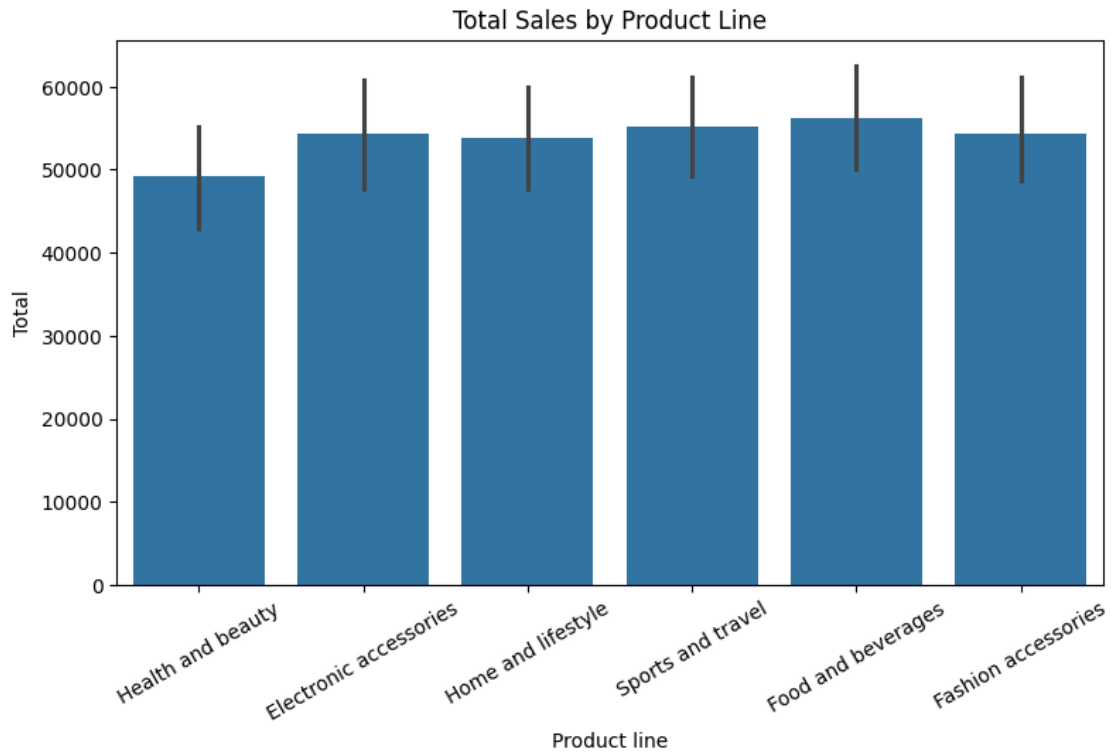
```
[51]: plt.figure(figsize=(7,5))
sns.barplot(x='TimeOfDay', y='Total',data=df,
            estimator='sum',order=['Morning','Afternoon','Evening','Night'])
plt.title("SALES PER TIME OF DAY")
plt.show()
```



2 Isesengura rya Product line

3 Ibi byerekana product line iri hejuru mu mafaranga yinjiye

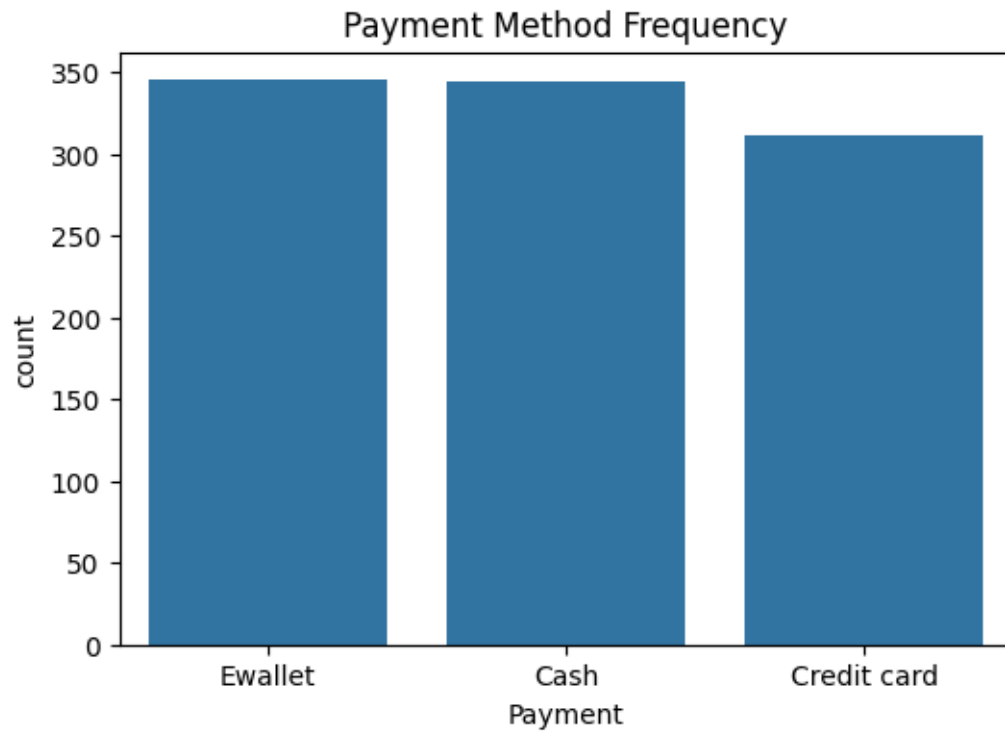
```
[52]: plt.figure(figsize=(9,5))
sns.barplot(x='Product line', y='Total', data=df, estimator='sum')
plt.title("Total Sales by Product Line")
plt.xticks(rotation=30)
plt.show()
```



4 Isesengura rya Payment

5 Ibi byerekana uburyo bwo kwishyura bukunze gukoreshwa (Cash, Card, E-wallet).

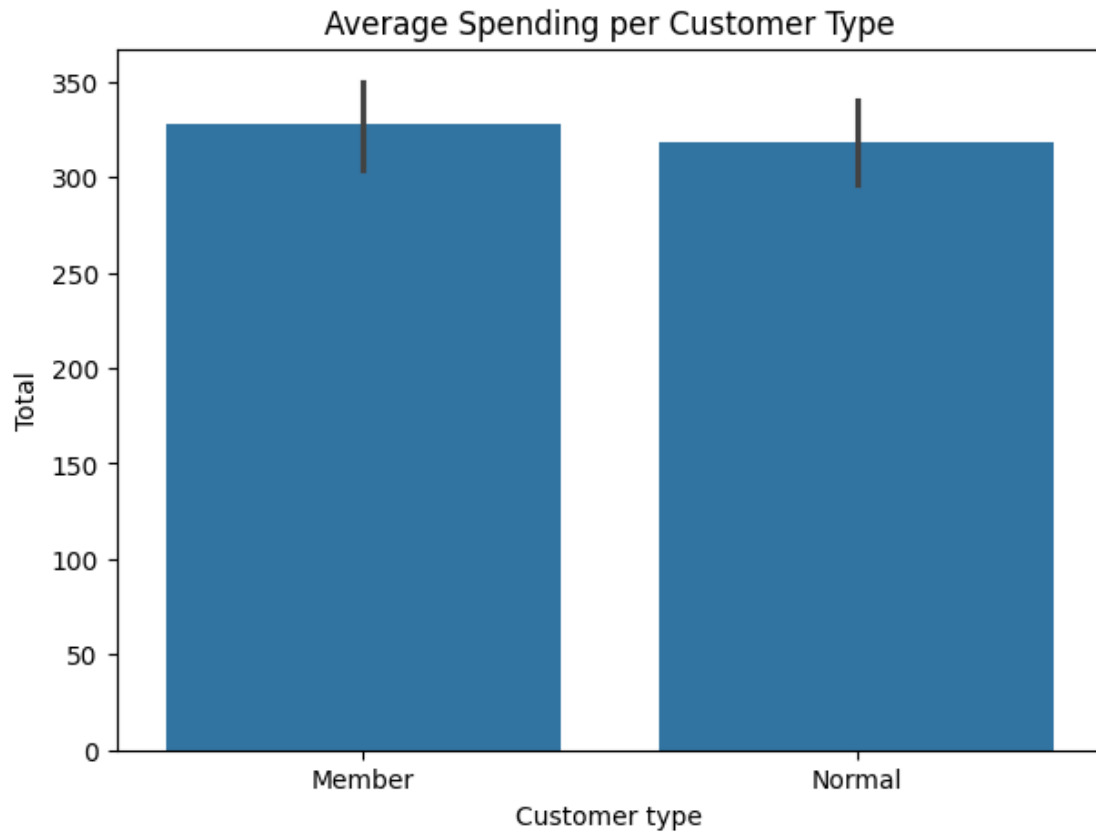
```
[54]: plt.figure(figsize=(6,4))
sns.countplot(x='Payment', data=df)
plt.title("Payment Method Frequency")
plt.show()
```



6 Isesengura rya Customer type

7 Ibi byerekana niba loyal customers batanga amafaranga menshi kurusha normal customers.

```
[55]: plt.figure(figsize=(7,5))
sns.barplot(x='Customer type', y='Total', data=df, estimator='mean')
plt.title("Average Spending per Customer Type")
plt.show()
```

8 Isesengura rya Rating

9 Ibi bitwereka niba abagura byinshi batanga rating nziza cyangwa mbi.

```
[56]: plt.figure(figsize=(7,5))
sns.scatterplot(x='Total', y='Rating', data=df, hue='Branch')
plt.title("Customer Rating vs Total Spending")
plt.show
```

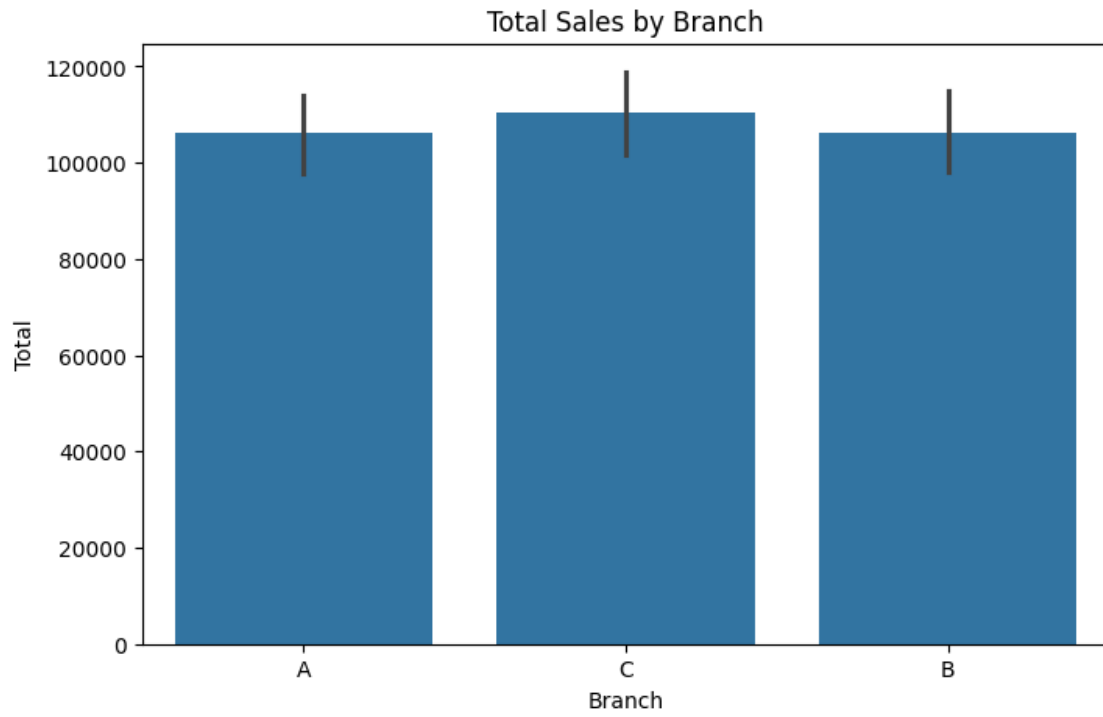
```
[56]: <function matplotlib.pyplot.show(close=None, block=None)>
```



10 Isesengura rya Branch/City

11 Bitwereka ishami (branch) rihiga ayandi mu kubyara inyungu.

```
[57]: plt.figure(figsize=(8,5))
sns.barplot(x='Branch', y='Total', data=df, estimator='sum')
plt.title("Total Sales by Branch")
plt.show()
```



[]: