

Quick access

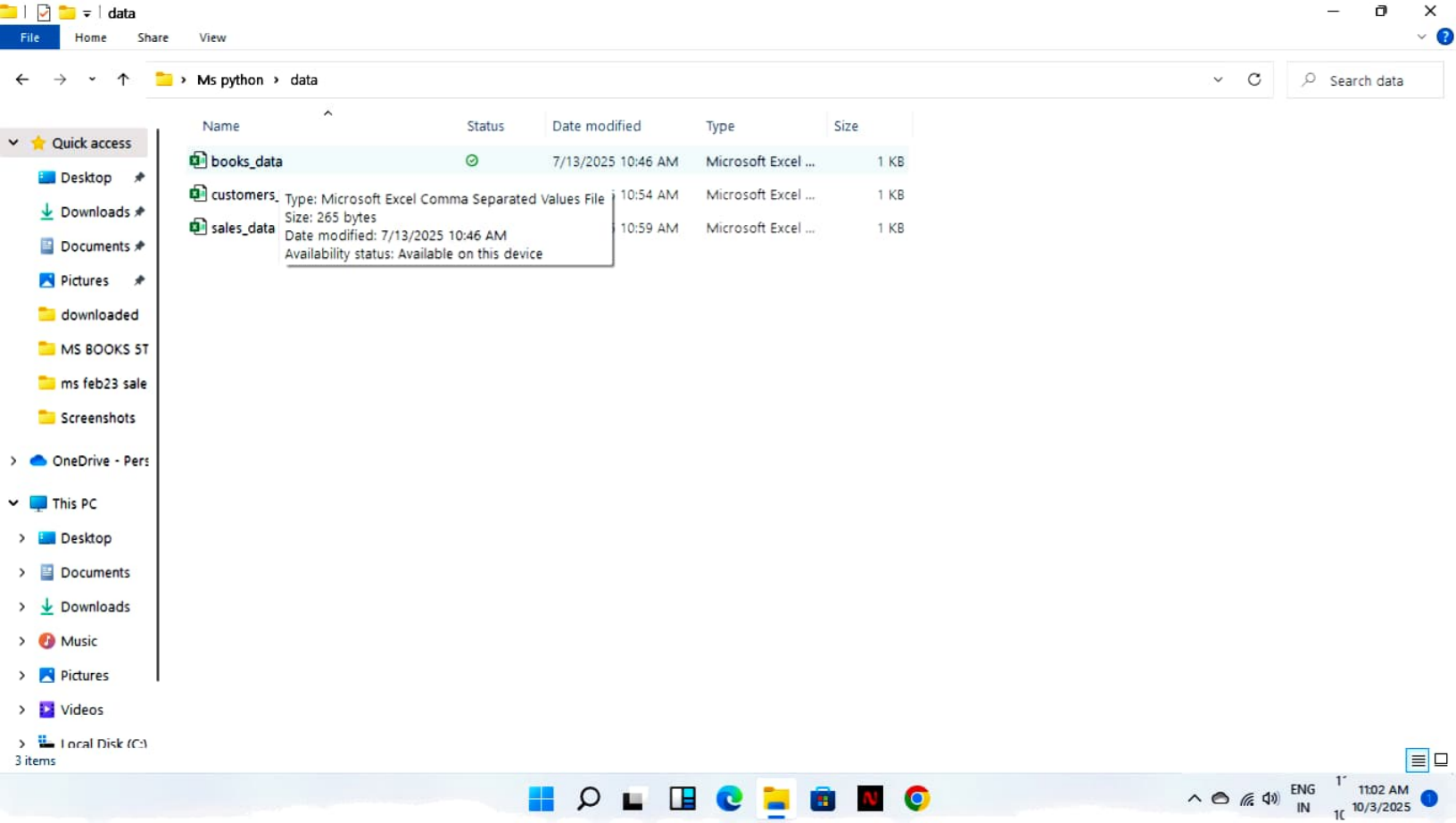
- Desktop
- Downloads
- Documents
- Pictures
- downloaded
- MS BOOKS ST
- ms feb23 sale
- Screenshots

OneDrive - Pers

This PC

- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- Local Disk (C:)

Name	Status	Date modified	Type	Size
data		7/13/2025 10:59 AM	File folder	
notebooks		7/16/2025 10:09 AM	File folder	
outputs		7/13/2025 11:25 AM	File folder	



Quick access

- Desktop
- Downloads
- Documents
- Pictures
- downloaded
- MS BOOKS ST
- ms feb23 sale
- Screenshots

OneDrive - Pers

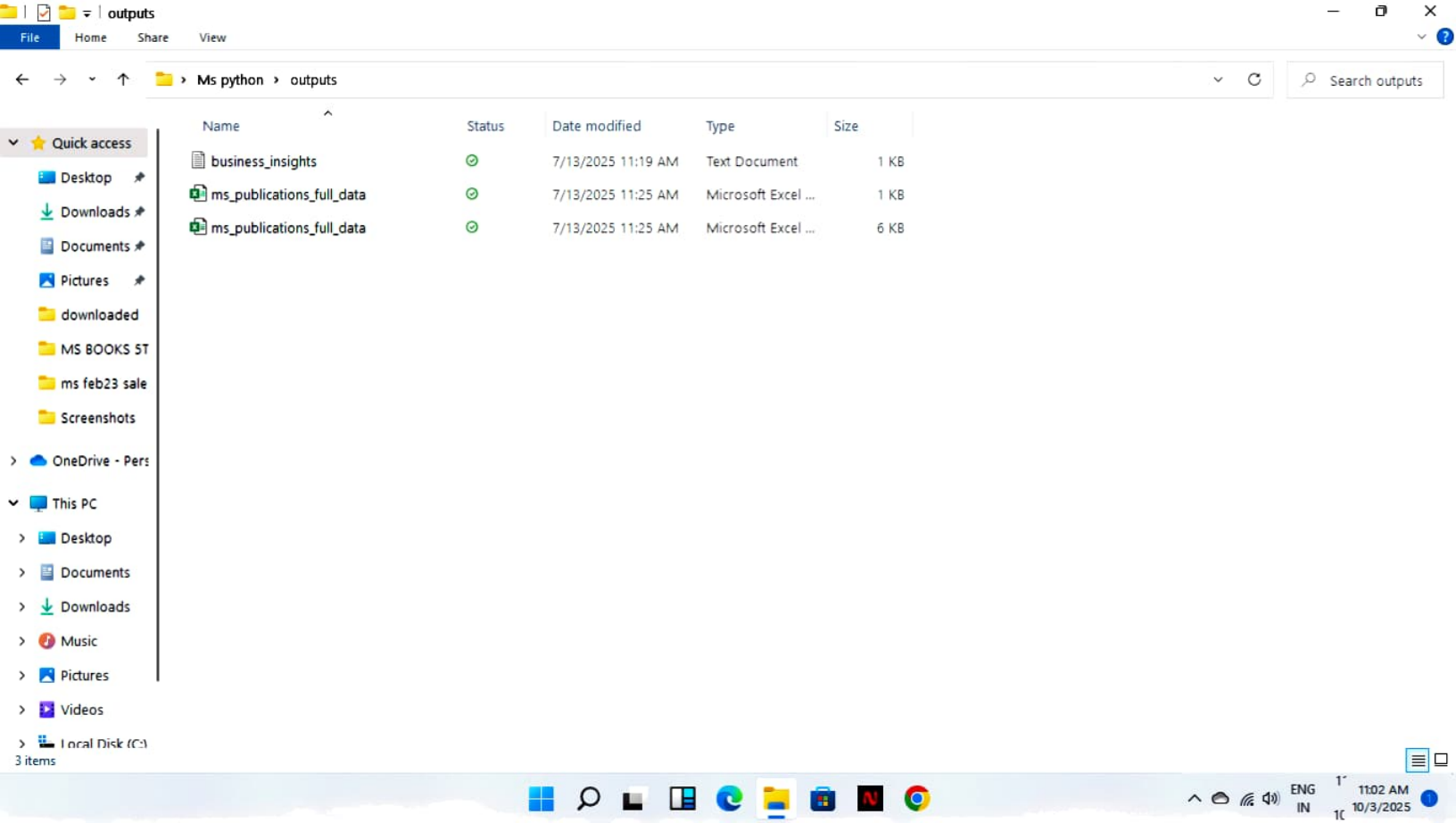
This PC

- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- Local Disk (C:)

3 items

Name	Status	Date modified	Type	Size
books_data		7/13/2025 10:46 AM	Microsoft Excel ...	1 KB
customers		10:54 AM	Microsoft Excel ...	1 KB
sales_data		10:59 AM	Microsoft Excel ...	1 KB

Type: Microsoft Excel Comma Separated Values File  
Size: 265 bytes  
Date modified: 7/13/2025 10:46 AM  
Availability status: Available on this device



Name	Status	Date modified	Type	Size
business_insights	✓	7/13/2025 11:19 AM	Text Document	1 KB
ms_publications_full_data	✓	7/13/2025 11:25 AM	Microsoft Excel ...	1 KB
ms_publications_full_data	✓	7/13/2025 11:25 AM	Microsoft Excel ...	6 KB

```
[notice] A new release of pip is available: 25.0.1 -> 25.1.1
```



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

%matplotlib inline

[3]: # Creating the Books dataset
books_data = pd.DataFrame({
    'book_id': [101, 102, 103, 104, 105],
    'title': ['Economics Basics', 'Advanced Physics', 'Political Science', 'Mathematics Core', 'English Fluency'],
    'subject': ['Economics', 'Physics', 'Pol Science', 'Mathematics', 'English'],
    'course': ['BA Prog', 'BSc Physics', 'BA Pol Sci', 'BSc Maths', 'BA Prog'],
    'price': [250, 400, 300, 350, 200]
})

# Display the dataset
books_data
# this command simply represents complete formatted table
```

[3]:	book_id	title	subject	course	price
0	101	Economics Basics	Economics	BA Prog	250

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

[3]:

	book_id	title	subject	course	price
0	101	Economics Basics	Economics	BA Prog	250
1	102	Advanced Physics	Physics	BSc Physics	400
2	103	Political Science	Pol Science	BA Pol Sci	300
3	104	Mathematics Core	Mathematics	BSc Maths	350
4	105	English Fluency	English	BA Prog	200

[4]:

```
# Save the dataset to a CSV file
books_data.to_csv('../data/books_data.csv', index=False)
```

[5]:

```
# Load books data from the CSV
books = pd.read_csv('../data/books_data.csv')

# Display loaded data
books
```

[5]:

	book_id	title	subject	course	price
0	101	Economics Basics	Economics	BA Prog	250

11:11 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

ms\_publications\_analysis Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

JupyterLab

Code

[5]:

	book_id	title	subject	course	price
0	101	Economics Basics	Economics	BA Prog	250
1	102	Advanced Physics	Physics	BSc Physics	400
2	103	Political Science	Pol Science	BA Pol Sci	300
3	104	Mathematics Core	Mathematics	BSc Maths	350
4	105	English Fluency	English	BA Prog	200

[6]:

```
# Creating the Customers dataset
customers_data = pd.DataFrame({
    'customer_id': [501, 502, 503, 504, 505],
    'name': ['Rahul Sharma', 'Delhi Book Depot', 'Priya Verma', 'NCR Book Store', 'Anjali Singh'],
    'location': ['Delhi', 'Delhi', 'Noida', 'Gurgaon', 'Faridabad'],
    'type': ['Student', 'Depot', 'Student', 'Depot', 'Student']
})

# Display the dataset
customers_data
```

[6]:

	customer_id	name	location	type
0	501	Rahul Sharma	Delhi	Student

11:11 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

Jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

JupyterLab

Python 3 (ipykernel)

[6]:

	customer_id	name	location	type
0	501	Rahul Sharma	Delhi	Student
1	502	Delhi Book Depot	Delhi	Depot
2	503	Priya Verma	Noida	Student
3	504	NCR Book Store	Gurgaon	Depot
4	505	Anjali Singh	Faridabad	Student

[7]:

```
# Save to CSV
customers_data.to_csv('../data/customers_data.csv', index=False)
```

[8]:

```
# Load customers data
customers = pd.read_csv('../data/customers_data.csv')

# Display Loaded data
customers
```

[8]:

	customer_id	name	location	type
0	501	Rahul Sharma	Delhi	Student

11:12 AM

10/3/2025



ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

Jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

[8]:

	customer_id	name	location	type
0	501	Rahul Sharma	Delhi	Student
1	502	Delhi Book Depot	Delhi	Depot
2	503	Priya Verma	Noida	Student
3	504	NCR Book Store	Gurgaon	Depot
4	505	Anjali Singh	Faridabad	Student

[9]:

```
# Creating the Sales dataset
sales_data = pd.DataFrame({
    'sale_id': [1001, 1002, 1003, 1004, 1005],
    'date': ['2024-01-05', '2024-01-12', '2024-02-10', '2024-02-20', '2024-03-01'],
    'book_id': [101, 102, 103, 104, 105],
    'customer_id': [501, 502, 503, 504, 505],
    'quantity': [2, 10, 1, 5, 3],
    'total_price': [500, 4000, 300, 1750, 600]
})

# Display the sales dataset
sales_data
```

11

11:12 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

JupyterLab

Python 3 (ipykernel)

[9]:

	sale_id	date	book_id	customer_id	quantity	total_price
0	1001	2024-01-05	101	501	2	500
1	1002	2024-01-12	102	502	10	4000
2	1003	2024-02-10	103	503	1	300
3	1004	2024-02-20	104	504	5	1750
4	1005	2024-03-01	105	505	3	600

[10]:

```
# Save the sales dataset to CSV
sales_data.to_csv('../data/sales_data.csv', index=False)
```

[11]:

```
# Load sales data
sales = pd.read_csv('../data/sales_data.csv')

# Display loaded sales data
sales
```

[11]:

	sale_id	date	book_id	customer_id	quantity	total_price
0	1001	2024-01-05	101	501	2	500

11:12 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

+

✂

📄

▶

■

🔄

▶▶

Code

▼

Not Trusted

JupyterLab

Python 3 (ipykernel)

[11]:

	sale_id	date	book_id	customer_id	quantity	total_price
0	1001	2024-01-05	101	501	2	500
1	1002	2024-01-12	102	502	10	4000
2	1003	2024-02-10	103	503	1	300
3	1004	2024-02-20	104	504	5	1750
4	1005	2024-03-01	105	505	3	600

[12]:

# View first few records of books

print("📖 Books Data: ")

print(books.head())

# View first few records of customers

print("\n👤 Customers Data: ")

print(customers.head())

# View first few records of sales

print("\n📄 Sales Data: ")

print(sales.head())

📖 Books Data:

book_id	title	subject	course	price
---------	-------	---------	--------	-------

Windows Taskbar

System Tray

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

ms\_publications\_analysis Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Books Data:

	book_id	title	subject	course	price
0	101	Economics Basics	Economics	BA Prog	250
1	102	Advanced Physics	Physics	BSc Physics	400
2	103	Political Science	Pol Science	BA Pol Sci	300
3	104	Mathematics Core	Mathematics	BSc Maths	350
4	105	English Fluency	English	BA Prog	200

Customers Data:

	customer_id	name	location	type
0	501	Rahul Sharma	Delhi	Student
1	502	Delhi Book Depot	Delhi	Depot
2	503	Priya Verma	Noida	Student
3	504	NCR Book Store	Gurgaon	Depot
4	505	Anjali Singh	Faridabad	Student

Sales Data:

	sale_id	date	book_id	customer_id	quantity	total_price
0	1001	2024-01-05	101	501	2	500
1	1002	2024-01-12	102	502	10	4000
2	1003	2024-02-10	103	503	1	300
3	1004	2024-02-20	104	504	5	1750
4	1005	2024-03-01	105	505	3	600

[13]: print("Books Info:")  
print(books.info())

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Code

[13]:

```
print("Books Info:")
print(books.info())

print("\nCustomers Info:")
print(customers.info())

print("\nSales Info:")
print(sales.info())
```

Books Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   book_id     5 non-null      int64
 1   title       5 non-null      object
 2   subject     5 non-null      object
 3   course      5 non-null      object
 4   price       5 non-null      int64
dtypes: int64(2), object(3)
memory usage: 332.0+ bytes
None
```

Customers Info:

```
<class 'pandas.core.frame.DataFrame'>
```

11

11:12 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

Jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

```
[17]: # Merge sales with books on 'book_id'
sales_books = pd.merge(sales, books, on='book_id', how='left')

# Merge the result with customers on 'customer_id'
full_data = pd.merge(sales_books, customers, on='customer_id', how='left')

# View combined data
full_data
```

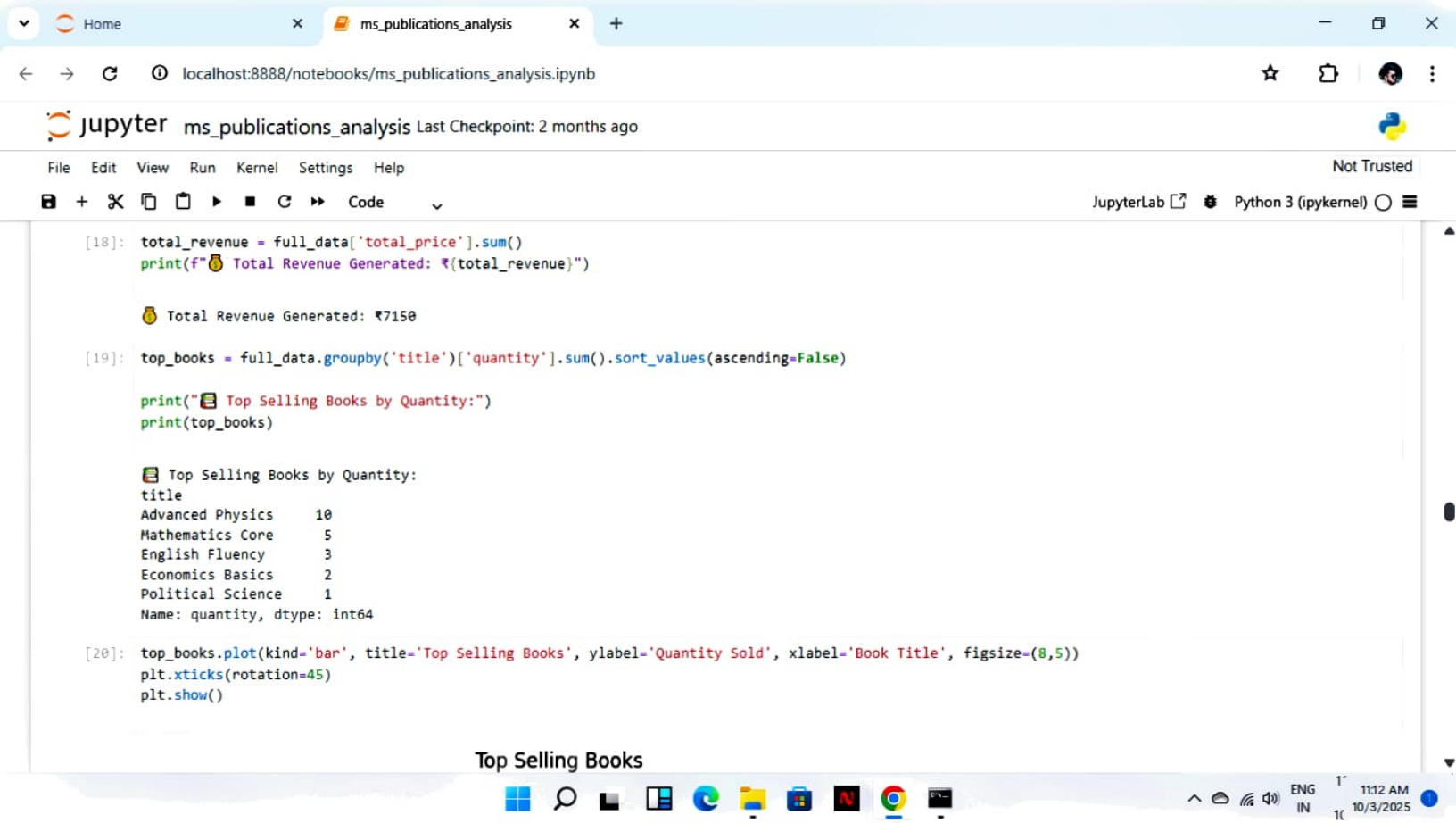
	sale_id	date	book_id	customer_id	quantity	total_price	title	subject	course	price	name	location	type
0	1001	2024-01-05	101	501	2	500	Economics Basics	Economics	BA Prog	250	Rahul Sharma	Delhi	Student
1	1002	2024-01-12	102	502	10	4000	Advanced Physics	Physics	BSc Physics	400	Delhi Book Depot	Delhi	Depot
2	1003	2024-02-10	103	503	1	300	Political Science	Pol Science	BA Pol Sci	300	Priya Verma	Noida	Student
3	1004	2024-02-20	104	504	5	1750	Mathematics Core	Mathematics	BSc Maths	350	NCR Book Store	Gurgaon	Depot
4	1005	2024-03-01	105	505	3	600	English Fluency	English	BA Prog	200	Anjali Singh	Faridabad	Student

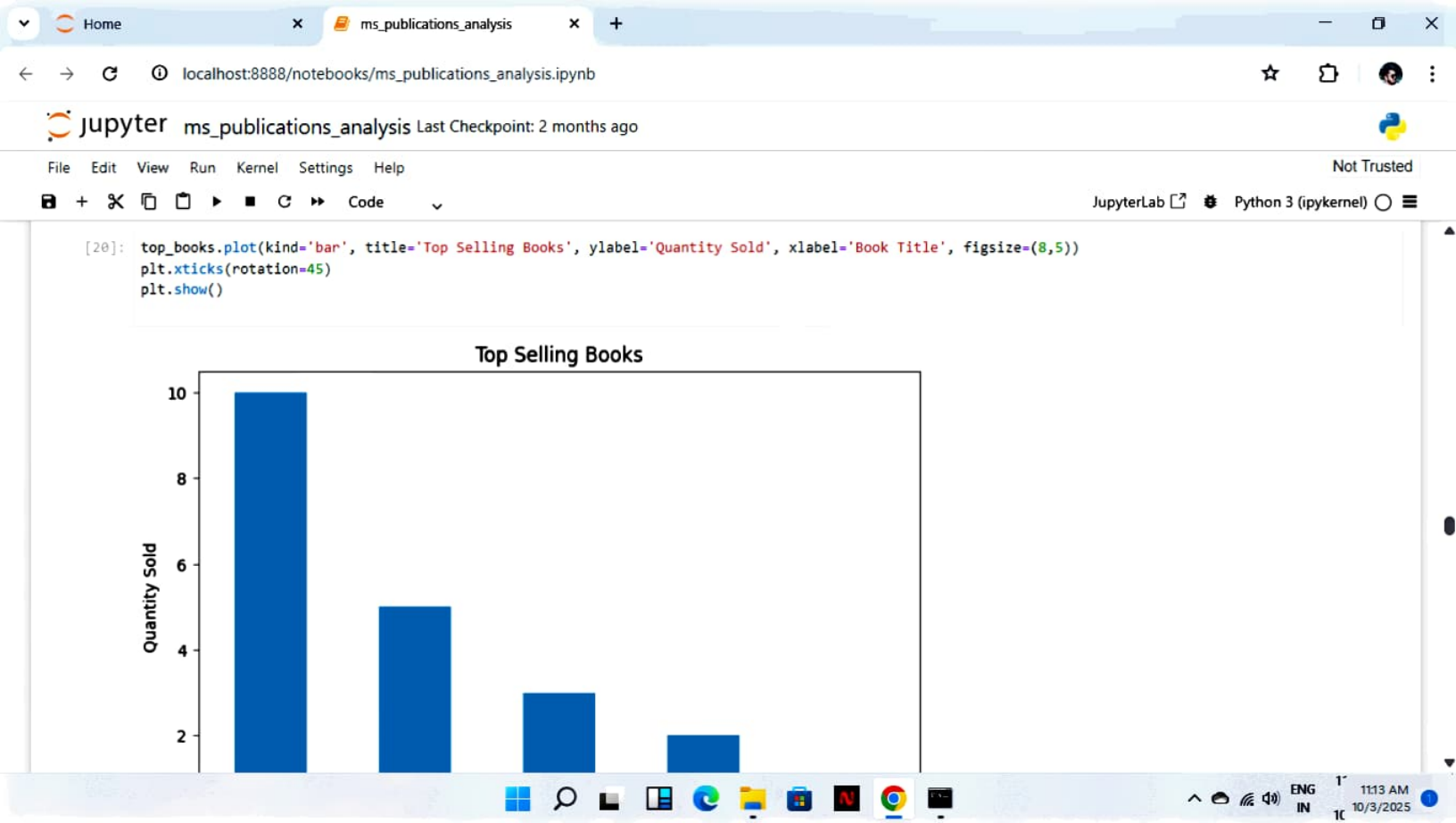
```
[18]: total_revenue = full_data['total_price'].sum()
print(f"👉 Total Revenue Generated: ₹{total_revenue}")
```

11

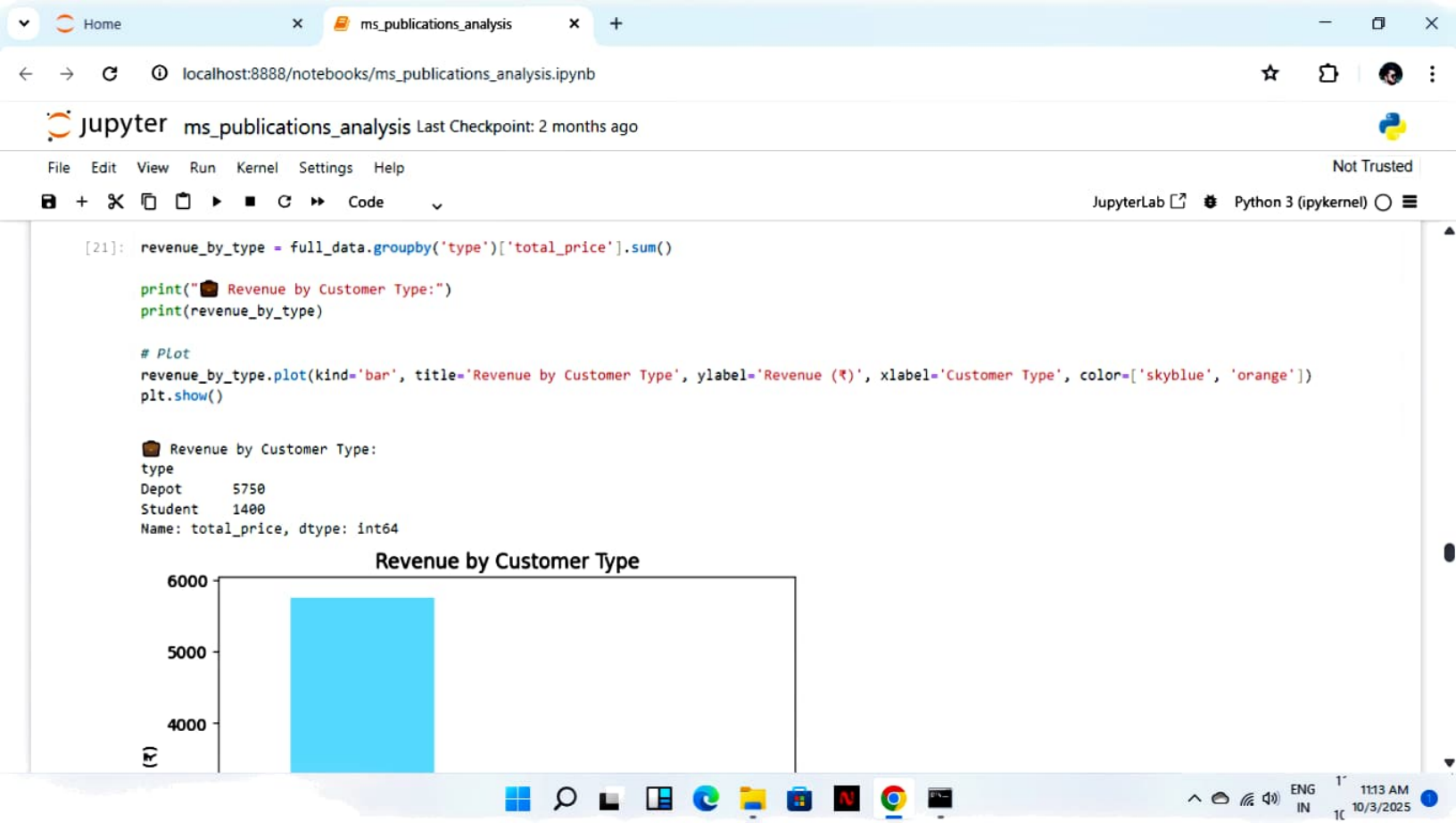
11:12 AM

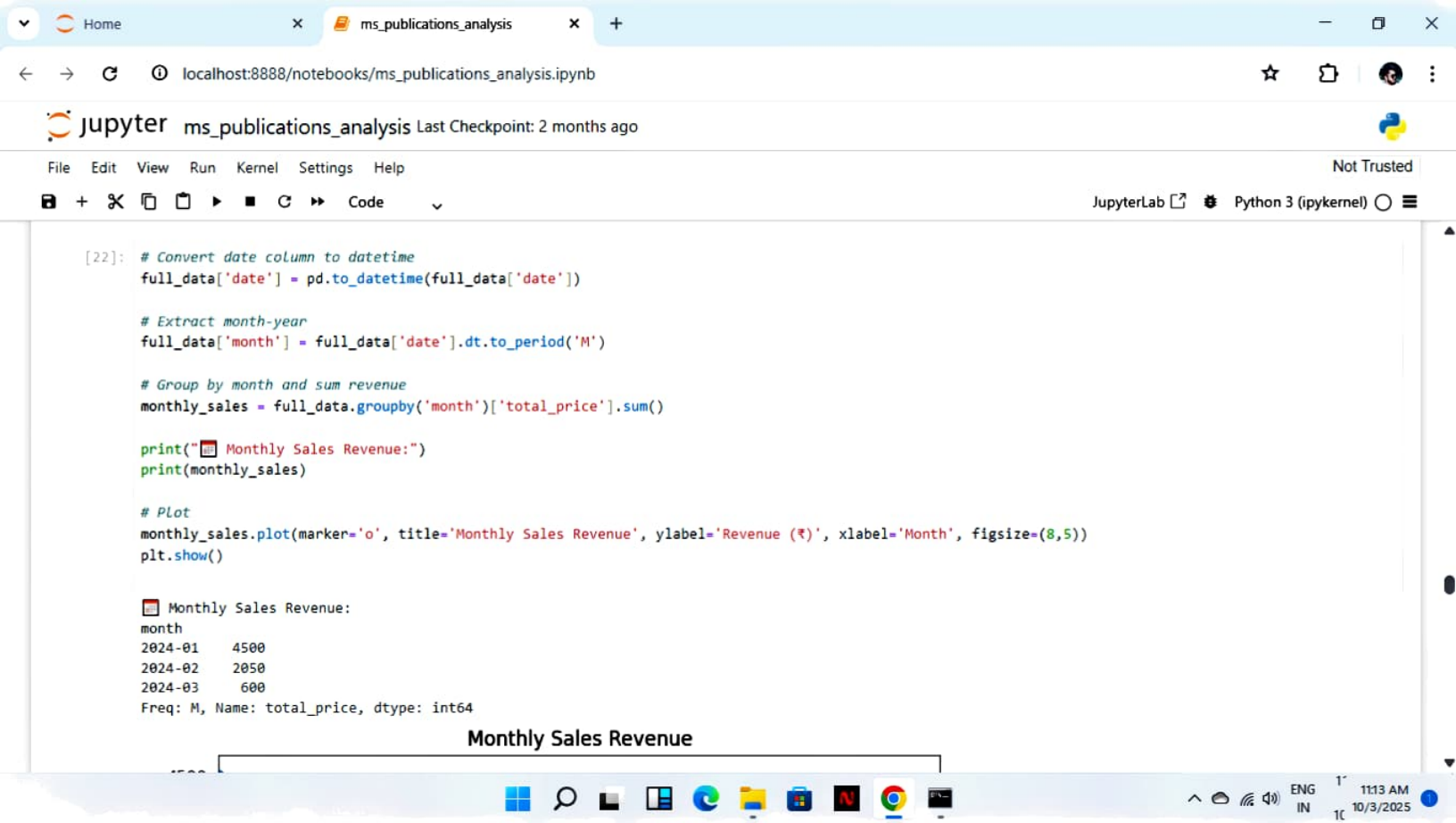
10/3/2025

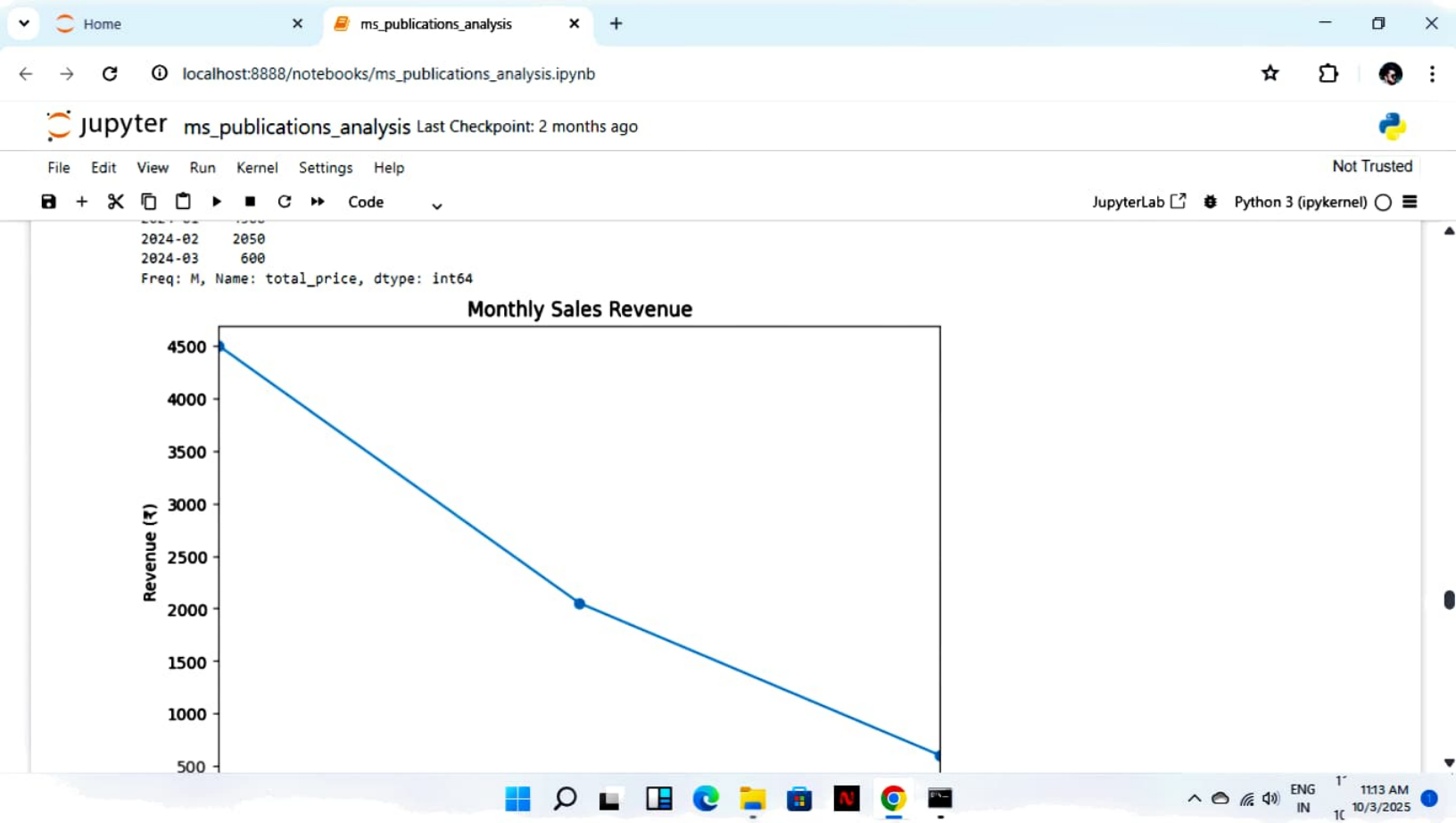












ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

Jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Month

```
[23]: print("🔥 BUSINESS INSIGHTS & RECOMMENDATIONS")
      print("-" * 50)

      print(f"1. Total Revenue generated: ₹{total_revenue}\n")

      print("2. Top-Selling Books by Quantity:")
      print(top_books, "\n")

      print("3. Revenue by Customer Type:")
      print(revenue_by_type, "\n")

      top_type = revenue_by_type.idxmax()
      print(f"📦 Recommendation: Focus marketing efforts more on '{top_type}' customers as they contribute the highest revenue.\n")

      print("4. Monthly Sales Trend:")
      print(monthly_sales, "\n")

      best_month = monthly_sales.idxmax()
      print(f"📦 Recommendation: Increase stock and promotions around {best_month} as it had the highest sales.\n")

      🔥 BUSINESS INSIGHTS & RECOMMENDATIONS
      -----
      1. Total Revenue generated: ₹7150
```

11:13 AM

10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

+

+

✂

📄

📄

▶

■

🔄

▶▶

Code

▼

Not Trusted

JupyterLab

Python 3 (ipykernel)

```
print(f"📌 Recommendation: Increase stock and promotions around {best_month} as it had the highest sales.\n")
```

📌 BUSINESS INSIGHTS & RECOMMENDATIONS

-----

1. Total Revenue generated: ₹7150

2. Top-Selling Books by Quantity:

title

Advanced Physics 10

Mathematics Core 5

English Fluency 3

Economics Basics 2

Political Science 1

Name: quantity, dtype: int64

3. Revenue by Customer Type:

type

Depot 5750

Student 1400

Name: total\_price, dtype: int64

📌 Recommendation: Focus marketing efforts more on 'Depot' customers as they contribute the highest revenue.

4. Monthly Sales Trend:

month

2024-01 4500

2024-02 2050

2024-03 6000

🪟 🔍 🖨 📄 🌐 📁 🛒 📺

ENG IN 11:13 AM 10/3/2025

ms\_publications\_analysis

localhost:8888/notebooks/ms\_publications\_analysis.ipynb

jupyter

ms\_publications\_analysis

Last Checkpoint: 2 months ago

File Edit View Run Kernel Settings Help

Python 3 (ipykernel)

Not Trusted

Code

```
f.write(f"Recommendation: Peak sales occurred in {best_month}. Plan for stock and promotions accordingly.\n")
```

```
[28]: # Export the full merged data to Excel
full_data.to_excel('../outputs/ms_publications_full_data.xlsx', index=False)

print("✅ Full data exported successfully to outputs/ms_publications_full_data.xlsx")

✅ Full data exported successfully to outputs/ms_publications_full_data.xlsx

[29]: # Export to CSV
full_data.to_csv('../outputs/ms_publications_full_data.csv', index=False)

print("✅ Full data exported successfully to outputs/ms_publications_full_data.csv")

✅ Full data exported successfully to outputs/ms_publications_full_data.csv

[2]: # Pie chart of sales distribution by book title
sales_by_book = full_data.groupby('title')['quantity'].sum()
plt.figure(figsize=(7,7))
sales_by_book.plot.pie(autopct='%1.1f%%', startangle=90, cmap='Set3', ylabel='')
plt.title('Sales Distribution by Book Title')
plt.ylabel('')
plt.show()
```

11:14 AM

10/3/2025