

AZURE PROJECT

MI 462

Joo Yong Yoon

A decorative graphic on the left side of the slide, composed of several overlapping geometric shapes and patterns. It includes a blue triangle with white diagonal lines, a light blue circle, a dark blue square with concentric circles, a dark purple triangle, a blue square, a grey square with concentric lines, a bright pink square with white concentric semi-circles, and a dark purple square at the bottom. A small blue circle is positioned at the intersection of the dark purple triangle and the bright pink square.

CHOICE OF AZURE SERVICES

- **Visual Studio (using Python for visualization)**
- **Power BI**



GOAL

E-COMMERCE

Using the historical data, understand the effective ecommerce strategy. Make predictions based on the findings.

INDIAN MARKET

Analyze Indian cosmetics market. Analyze historical trend & see what affect the net sales the most.

File Edit Selection View Go Run Terminal Help workspace (Workspace)

EXPLORER

OPEN EDITORS 1 unsaved

- import numpy as np Untitled-1
- Untitled-1.ipynb C:\Users\jyy21\...
- 2019-Dec.csv C:\Users\jyy21\Des...
- 2019-Nov.csv C:\Users\jyy21\Des...
- 2019-Oct.csv C:\Users\jyy21\Des...
- 2020-Feb.csv C:\Users\jyy21\Des...
- 2020-Jan.csv C:\Users\jyy21\Des...

NO FOLDER OPENED

OUTLINE

TIMELINE

import numpy as np
import pandas as pd

[2] ✓ 0.9s Python

pip install numpy

[1] Python

pip install pandas

[1] Python

```
df_19_oct=pd.read_csv(r'C:/Users/jyy21/Desktop/archive/2019-Oct.csv')
df_19_nov=pd.read_csv(r'C:/Users/jyy21/Desktop/archive/2019-Nov.csv')
df_19_dec=pd.read_csv(r'C:/Users/jyy21/Desktop/archive/2019-Dec.csv')
df_20_jan=pd.read_csv(r'C:/Users/jyy21/Desktop/archive/2020-Jan.csv')
df_20_feb=pd.read_csv(r'C:/Users/jyy21/Desktop/archive/2020-Feb.csv')
print(df_19_oct)
```

[6] ✓ 38.3s Python

...	event_time	event_type	product_id	\
0	2019-10-01 00:00:00 UTC	cart	5773203	
1	2019-10-01 00:00:03 UTC	cart	5773353	
2	2019-10-01 00:00:07 UTC	cart	5881589	
3	2019-10-01 00:00:07 UTC	cart	5723490	
4	2019-10-01 00:00:15 UTC	cart	5881449	
...	
4102278	2019-10-31 23:59:25 UTC	remove_from_cart	5848062	
4102279	2019-10-31 23:59:30 UTC	view	5885631	
4102280	2019-10-31 23:59:34 UTC	view	5784898	
4102281	2019-10-31 23:59:36 UTC	view	5870838	
4102282	2019-10-31 23:59:54 UTC	view	59996	

3.6.8 64-bit Cell 2 of 19

FileEditSelectionViewGoRunTerminalHelp

workspace (Workspace)

import numpy as npUntitled-1Untitled-1.ipynb2019-Dec.csv2019-Nov.csv2019-Oct.csv2020-Feb.csv2020-Jan.csv

OPEN EDITORS1 unsaved

import numpy as npUntitled-1

Untitled-1.ipynbC:\Users\jyy21\...

2019-Dec.csvC:\Users\jyy21\Des...

2019-Nov.csvC:\Users\jyy21\Des...

2019-Oct.csvC:\Users\jyy21\Des...

2020-Feb.csvC:\Users\jyy21\Des...

2020-Jan.csvC:\Users\jyy21\Des...

NO FOLDER OPENED

OUTLINE

TIMELINE

C: > Users > jyy21 > Desktop > Untitled-1.ipynb > pip install missingno

+ Code + Markdown | Run All | Clear All Outputs | Outline ...

Python 3.6.8

	category_id	category_code	brand	price	user_id	\
0	1487580005134238553	NaN	runail	2.62	463240011	
1	1487580005134238553	NaN	runail	2.62	463240011	
2	2151191071051219817	NaN	lovely	13.48	429681830	
3	1487580005134238553	NaN	runail	2.62	463240011	
4	1487580013522845895	NaN	lovely	0.56	429681830	
...	
4102278	1526733091857498510	NaN	NaN	9.21	429913900	
4102279	1783999064103190764	NaN	NaN	15.48	556138645	
4102280	1487580009143992338	NaN	NaN	1.78	564441099	
4102281	1487580007675986893	NaN	milv	0.79	429913900	
4102282	1487580012927254698	NaN	NaN	4.13	566280798	
...						
4102281	2f0bff3c-252f-4fe6-afcd-5d8a6a92839a					
4102282	4eaf7c71-e8bf-4ea6-b719-eca144466eb3					

[4102283 rows x 9 columns]

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

pip install missingno

[]

Python

import missingno as msno

[11] ✓ 6.1s

Python

all_data = pd.concat([df_19_oct,df_19_nov,df_19_dec,df_20_jan,df_20_feb])
del df_19_oct,df_19_nov,df_19_dec,df_20_jan,df_20_feb
all_data.shape

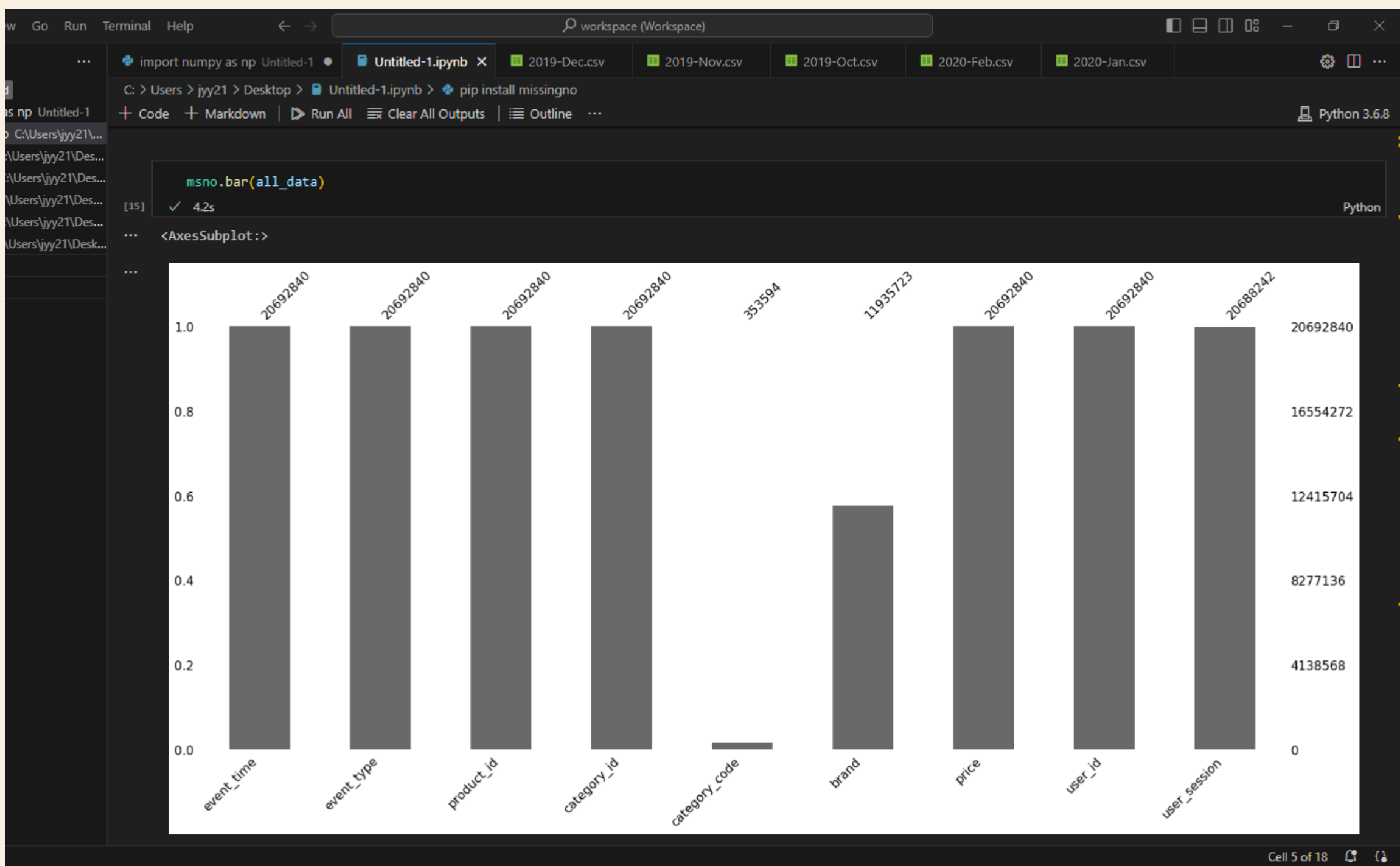
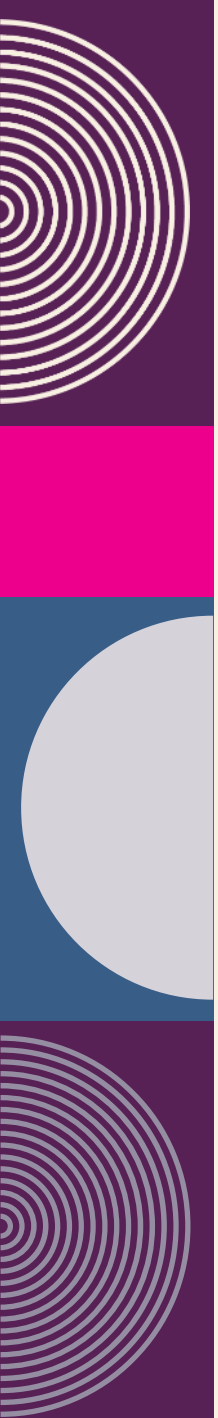
[14] ✓ 20.5s

Python

...

(20692840, 9)

13 20 03.6.8 64-bitCell 5 of 18



File Edit Selection View Go Run Terminal Help

workspace (Workspace)

import numpy as np Untitled-1

Untitled-1.ipynb X

2019-Dec.csv

2019-Nov.csv

2019-Oct.csv

2020-Feb.csv

2020-Jan.csv

Python 3.6.8

EXPLORER

OPEN EDITORS 1 unsaved

import numpy as np Untitled-1

Untitled-1.ipynb C:\Users\jyy21\...

2019-Dec.csv C:\Users\jyy21\Des...

2019-Nov.csv C:\Users\jyy21\Des...

2019-Oct.csv C:\Users\jyy21\Des...

2020-Feb.csv C:\Users\jyy21\Des...

2020-Jan.csv C:\Users\jyy21\Des...

NO FOLDER OPENED

OUTLINE

TIMELINE

C: > Users > jyy21 > Desktop > Untitled-1.ipynb > pip install missingno

+ Code + Markdown Run All Clear All Outputs Outline

all_data.drop(["user_session","category_code","brand"], axis = 1, inplace=True)

all_data

[16] ✓ 24.2s Python

	event_time	event_type	product_id	category_id	price	user_id
0	2019-10-01 00:00:00 UTC	cart	5773203	1487580005134238553	2.62	463240011
1	2019-10-01 00:00:03 UTC	cart	5773353	1487580005134238553	2.62	463240011
2	2019-10-01 00:00:07 UTC	cart	5881589	2151191071051219817	13.48	429681830
3	2019-10-01 00:00:07 UTC	cart	5723490	1487580005134238553	2.62	463240011
4	2019-10-01 00:00:15 UTC	cart	5881449	1487580013522845895	0.56	429681830
...
4156677	2020-02-29 23:59:32 UTC	view	5885416	1487580005092295511	6.27	622082947
4156678	2020-02-29 23:59:39 UTC	cart	5550686	1487580008145748965	1.11	459705611
4156679	2020-02-29 23:59:45 UTC	view	5850628	1602943681873052386	5.24	622090043
4156680	2020-02-29 23:59:54 UTC	view	5716351	1487580010872045658	0.79	619841242
4156681	2020-02-29 23:59:59 UTC	view	5927709	1487580013950664926	3.17	621669991

20692840 rows x 6 columns

all_data['event_time'] = all_data['event_time'].astype('datetime64')

[23] ✓ 22m 6.4s Python

```
all_data['event_date'] = all_data['event_time'].dt.date
all_data['event_hour'] = all_data['event_time'].dt.hour
all_data['Day'] = all_data['event_time'].dt.day_name()
```

13 20 0

Cell 5 of 18

File Edit Selection View Go Run Terminal Help workspace (Workspace)

EXPLORER

OPEN EDITORS 1 unsaved

- import numpy as np Untitled-1
- Untitled-1.ipynb C:\Users\jyy21\...
- 2019-Dec.csv C:\Users\jyy21\Des...
- 2019-Nov.csv C:\Users\jyy21\Des...
- 2019-Oct.csv C:\Users\jyy21\Des...
- 2020-Feb.csv C:\Users\jyy21\Des...
- 2020-Jan.csv C:\Users\jyy21\Des...

NO FOLDER OPENED

OUTLINE

TIMELINE

import numpy as np Untitled-1

C: > Users > jyy21 > Desktop > Untitled-1.ipynb > pip install missingno

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Python 3.6.8

```
[23] ✓ 22m 6.4s
```

```
all_data['event_time'] = all_data['event_time'].astype('datetime64')
```

```
[25] ✓ 27.1s
```

```
all_data['event_date'] = all_data['event_time'].dt.date
all_data['event_hour'] = all_data['event_time'].dt.hour
all_data['Day'] = all_data['event_time'].dt.day_name()
all_data['Month'] = all_data['event_time'].dt.month_name()
all_data['serial_number'] = np.arange(all_data.shape[0])
all_data['users'] = all_data['serial_number'].diff()
all_data.at[0, 'users'] = 1

all_data.head(5)
```

```
...
```

	event_time	event_type	product_id	category_id	price	user_id	event_date	event_hour	Day	Month	serial_number	users
0	2019-10-01 00:00:00	cart	5773203	1487580005134238553	2.62	463240011	2019-10-01	0	Tuesday	October	0	1.0
1	2019-10-01 00:00:03	cart	5773353	1487580005134238553	2.62	463240011	2019-10-01	0	Tuesday	October	1	1.0
2	2019-10-01 00:00:07	cart	5881589	2151191071051219817	13.48	429681830	2019-10-01	0	Tuesday	October	2	1.0
3	2019-10-01 00:00:07	cart	5723490	1487580005134238553	2.62	463240011	2019-10-01	0	Tuesday	October	3	1.0
4	2019-10-01 00:00:15	cart	5881449	1487580013522845895	0.56	429681830	2019-10-01	0	Tuesday	October	4	1.0

```
[ ]
```

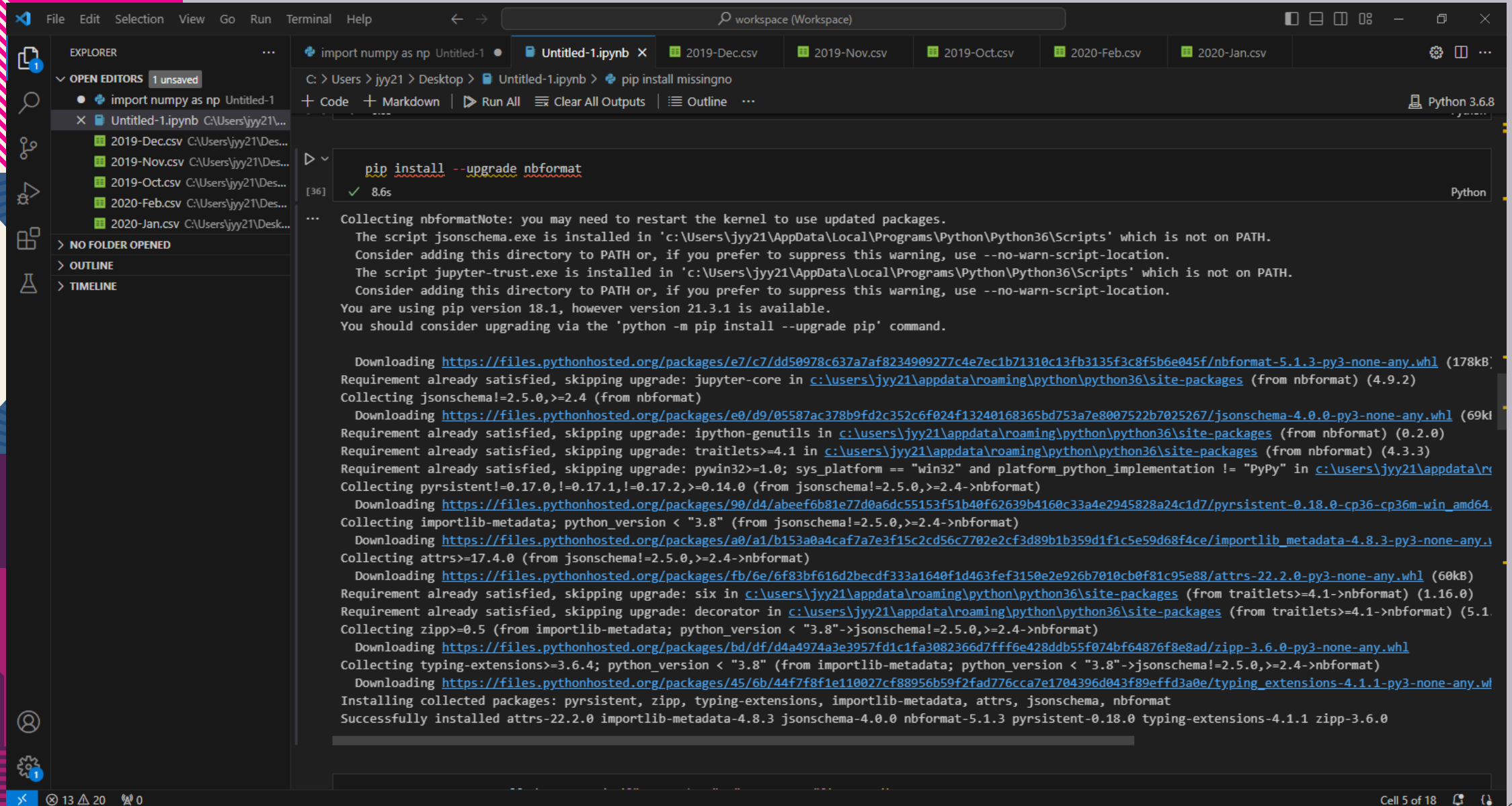
```
pip install plotly.express
```

```
[33] ✓ 0.0s
```

```
import plotly.express as px
```

Cell 5 of 18

ValueError: Mime type rendering requires nbformat>=4.2.0 but it is not installed



The screenshot shows the VS Code interface with a Jupyter Notebook. The Explorer pane on the left shows a workspace with several files, including 'Untitled-1.ipynb'. The main editor area displays the notebook content, which includes a terminal output for the command `pip install --upgrade nbformat`. The output shows the installation progress, including downloading packages and their dependencies. The status bar at the bottom indicates 'Cell 5 of 18'.

```
import numpy as np
Untitled-1
Untitled-1.ipynb
2019-Dec.csv
2019-Nov.csv
2019-Oct.csv
2020-Feb.csv
2020-Jan.csv
```

workspace (Workspace)

Python 3.6.8

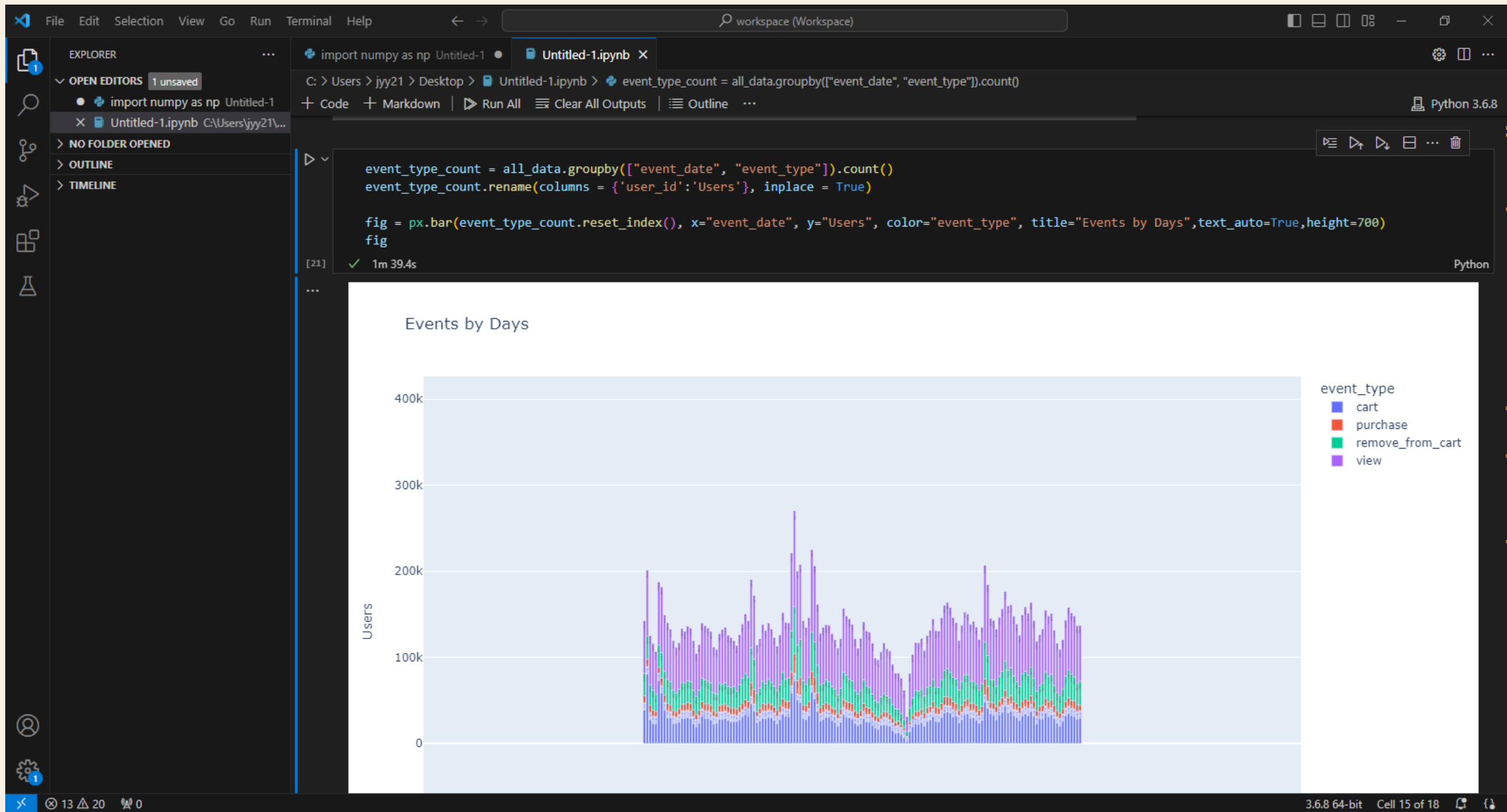
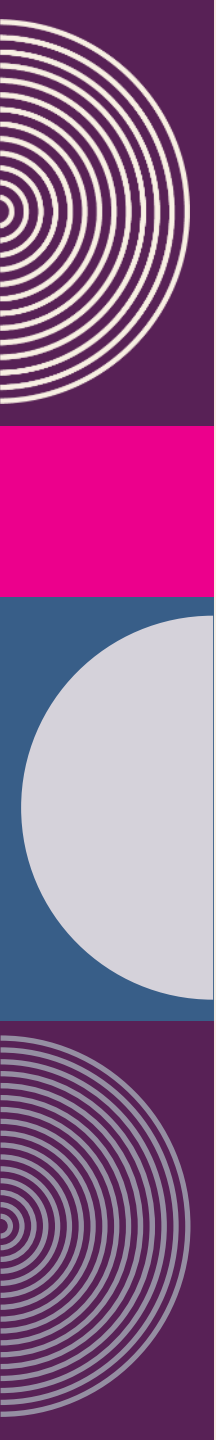
```
pip install --upgrade nbformat
```

[36] ✓ 8.6s

Collecting nbformatNote: you may need to restart the kernel to use updated packages.
The script jsonschema.exe is installed in 'c:\Users\jyy21\AppData\Local\Programs\Python\Python36\Scripts' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
The script jupyter-trust.exe is installed in 'c:\Users\jyy21\AppData\Local\Programs\Python\Python36\Scripts' which is not on PATH.
Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
You are using pip version 18.1, however version 21.3.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' command.

Downloading <https://files.pythonhosted.org/packages/e7/c7/dd50978c637a7af8234909277c4e7ec1b71310c13fb3135f3c8f5b6e045f/nbformat-5.1.3-py3-none-any.whl> (178kB)
Requirement already satisfied, skipping upgrade: jupyter-core in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from nbformat) (4.9.2)
Collecting jsonschema!=2.5.0,>=2.4 (from nbformat)
Downloading <https://files.pythonhosted.org/packages/e0/d9/05587ac378b9fd2c352c6f024f13240168365bd753a7e8007522b7025267/jsonschema-4.0.0-py3-none-any.whl> (69kB)
Requirement already satisfied, skipping upgrade: ipython-genutils in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from nbformat) (0.2.0)
Requirement already satisfied, skipping upgrade: traitlets>=4.1 in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from nbformat) (4.3.3)
Requirement already satisfied, skipping upgrade: pywin32>=1.0; sys_platform == "win32" and platform_python_implementation != "PyPy" in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from nbformat) (220.0.1)
Collecting pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 (from jsonschema!=2.5.0,>=2.4->nbformat)
Downloading https://files.pythonhosted.org/packages/90/d4/abeef6b81e77d0a6dc55153f51b40f62639b4160c33a4e2945828a24c1d7/pyrsistent-0.18.0-cp36-cp36m-win_amd64.whl (115kB)
Collecting importlib-metadata; python_version < "3.8" (from jsonschema!=2.5.0,>=2.4->nbformat)
Downloading https://files.pythonhosted.org/packages/a0/a1/b153a0a4caf7a7e3f15c2cd56c7702e2cf3d89b1b359d1f1c5e59d68f4ce/importlib_metadata-4.8.3-py3-none-any.whl (28kB)
Collecting attrs>=17.4.0 (from jsonschema!=2.5.0,>=2.4->nbformat)
Downloading <https://files.pythonhosted.org/packages/fb/6e/6f83bf616d2becdf333a1640f1d463fef3150e2e926b7010cb0f81c95e88/attrs-22.2.0-py3-none-any.whl> (60kB)
Requirement already satisfied, skipping upgrade: six in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from traitlets>=4.1->nbformat) (1.16.0)
Requirement already satisfied, skipping upgrade: decorator in c:\users\jyy21\appdata\roaming\python\python36\site-packages (from traitlets>=4.1->nbformat) (5.1.1)
Collecting zipp>=0.5 (from importlib-metadata; python_version < "3.8"->jsonschema!=2.5.0,>=2.4->nbformat)
Downloading <https://files.pythonhosted.org/packages/bd/df/d4a4974a3e3957fd1c1fa3082366d7fff6e428ddb55f074bf64876f8e8ad/zipp-3.6.0-py3-none-any.whl> (22kB)
Collecting typing-extensions>=3.6.4; python_version < "3.8" (from importlib-metadata; python_version < "3.8"->jsonschema!=2.5.0,>=2.4->nbformat)
Downloading https://files.pythonhosted.org/packages/45/6b/44f7f8f1e110027cf88956b59f2fad776cca7e1704396d043f89effd3a0e/typing_extensions-4.1.1-py3-none-any.whl (20kB)
Installing collected packages: pyrsistent, zipp, typing-extensions, importlib-metadata, attrs, jsonschema, nbformat
Successfully installed attrs-22.2.0 importlib-metadata-4.8.3 jsonschema-4.0.0 nbformat-5.1.3 pyrsistent-0.18.0 typing-extensions-4.1.1 zipp-3.6.0

Cell 5 of 18



> TIMELINE

Python 3.6.8

[48] \otimes 29.0s

```
ValueError: Mime type rendering requires nbformat>=4.2.0 but it is not installed
```

EXPLORER

OPEN EDITORS 1 unsaved

- import numpy as np Untitled-1
- Untitled-1.ipynb C:\Users\jyy21\...

NO FOLDER OPENED

OUTLINE

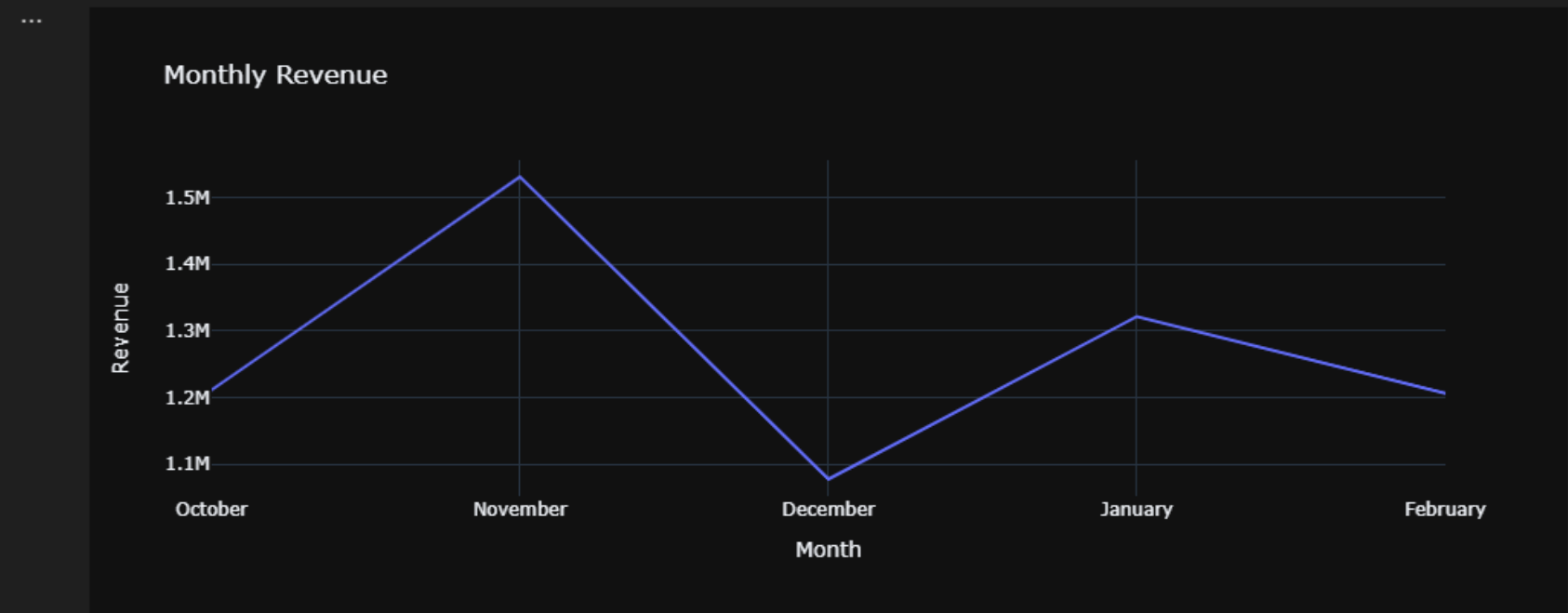
TIMELINE

1

```
import numpy as np Untitled-1 • Untitled-1.ipynb X
C: > Users > jyy21 > Desktop > Untitled-1.ipynb > all_data['event_time'] = all_data['event_time'].astype('datetime64')
+ Code + Markdown | Interrupt Clear All Outputs Go To | Outline ...
c:\Users\jyy21\AppData\Local\Programs\Python\Python36\lib\site-packages\plotly\basedatatypes.py in _ipython_display_(self)
830
831     if pio.renderers.render_on_display and pio.renderers.default:
--> 832         pio.show(self)
833     else:
834         print(repr(self))

c:\Users\jyy21\AppData\Local\Programs\Python\Python36\lib\site-packages\plotly\io\_renderers.py in show(fig, renderer, validate, **kwargs)
393     if not nbformat or Version(nbformat.__version__) < Version("4.2.0"):
394         raise ValueError(
--> 395             "Mime type rendering requires nbformat>=4.2.0 but it is not installed"
396         )
397

ValueError: Mime type rendering requires nbformat>=4.2.0 but it is not installed
```



File

Edit

Selection

View

Go

Run

Terminal

Help

workspace (Workspace)

EXPLORER

OPEN EDITORS 1 unsaved

import numpy as np Untitled-1

Untitled-1.ipynb C:\Users\jyy21\...

NO FOLDER OPENED

OUTLINE

TIMELINE

import numpy as np Untitled-1

Untitled-1.ipynb X

C: > Users > jyy21 > Desktop > Untitled-1.ipynb > all_data['event_time'] = all_data['event_time'].astype('datetime64')

+ Code + Markdown | Interrupt Clear All Outputs Go To | Outline ...

Python 3.6.8

event_type_count2 = all_data.groupby(["Day", "event_type"]).count()
event_type_count2.rename(columns = {'user_id': 'Users'}, inplace = True)

cat = ["5", "5", "5", "5", "1", "1", "1", "1", "6", "6", "6", "6", "7", "7", "7", "7", "4", "4", "4", "4", "2", "2", "2", "2", "3", "3", "3", "3"]
event_type_count2 = event_type_count2.assign(days=cat)
event_type_count2.sort_values(by=['days'], ascending=True, inplace=True)

event2 = df_purchase.groupby(["Day"]).sum()
cat3 = ["5", "1", "6", "7", "4", "2", "3"]
event2 = event2.assign(days=cat3)
event2.sort_values(by=['days'], ascending=True, inplace=True)

event_type_count3 = event_type_count2
event_type_count3.drop(["product_id", "product_id", "category_id", "event_date", "price", "event_time", "users", "serial_number", "days", "event_hour", "Month"], axis=1)
event_type_count3.rename(columns = {'user_id': 'Users'}, inplace = True)
event_type_count3

[51] ✓ 1m 12.2s

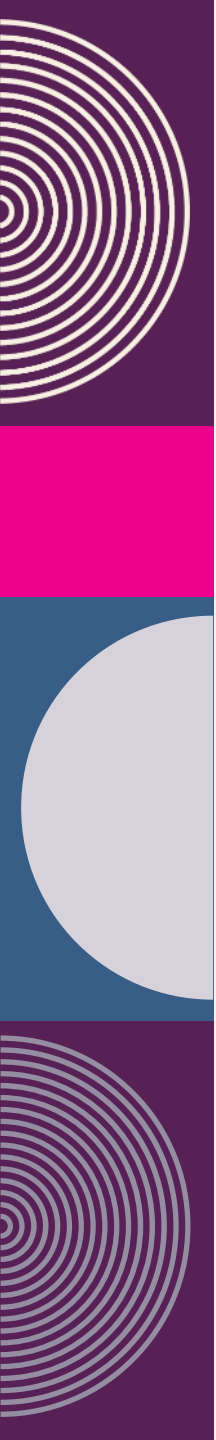
Python

...

		Users
Day	event_type	
Monday	cart	847239
	purchase	185220
	remove_from_cart	583607
	view	1410879
Tuesday	view	1421345
	remove_from_cart	585646
	purchase	192494
	cart	846961

13 20 0

Cell 10 of 18



File Edit Selection View Go Run Terminal Help

workspace (Workspace)

EXPLORER

OPEN EDITORS 1 unsaved

import numpy as np Untitled-1

Untitled-1.ipynb C:\Users\jyy21\...

NO FOLDER OPENED

OUTLINE

TIMELINE

import numpy as np Untitled-1

Untitled-1.ipynb X

C: > Users > jyy21 > Desktop > Untitled-1.ipynb > all_data['event_time'] = all_data['event_time'].astype('datetime64')

Code Markdown Interrupt Clear All Outputs Go To Outline

	remove_from_cart	583607
	view	1410879
Tuesday	view	1421345
	remove_from_cart	585646
	purchase	192494
	cart	846961
Wednesday	view	1449201
	purchase	184437
	cart	885990
	remove_from_cart	582926
Thursday	view	1453974
	remove_from_cart	606671
	purchase	206168
	cart	867441
Friday	cart	800277
	view	1377729
	remove_from_cart	566754
	purchase	198350
Saturday	view	1257366
	remove_from_cart	518252
	purchase	155781
	cart	730764
Sunday	remove_from_cart	535823
	cart	789661
	view	1287327
	purchase	164557

13 20 0

File

Home

Insert

Modeling

View

Optimize

Help

Paste

Cut

Copy

Format painter

Get data

Excel workbook

OneLake data hub

SQL Server

Enter data

Dataverse

Recent sources

Transform data

Refresh

New visual

Text box

More visuals

New measure

Quick measure

Sensitivity

Publish

Clipboard

Data

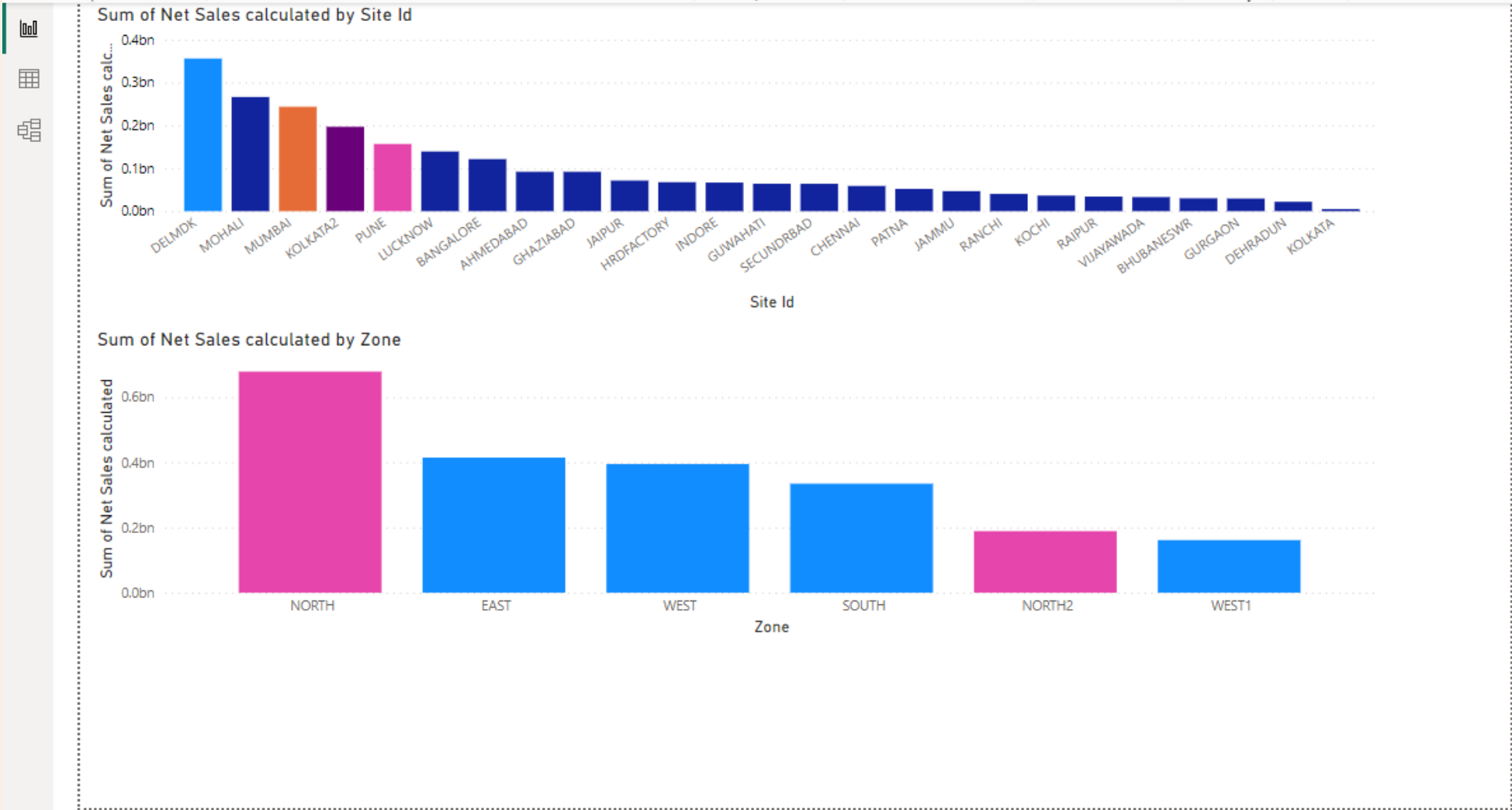
Queries

Insert

Calculations

Sensitivity

Share



Visualizations

Build visual

Filters

Values

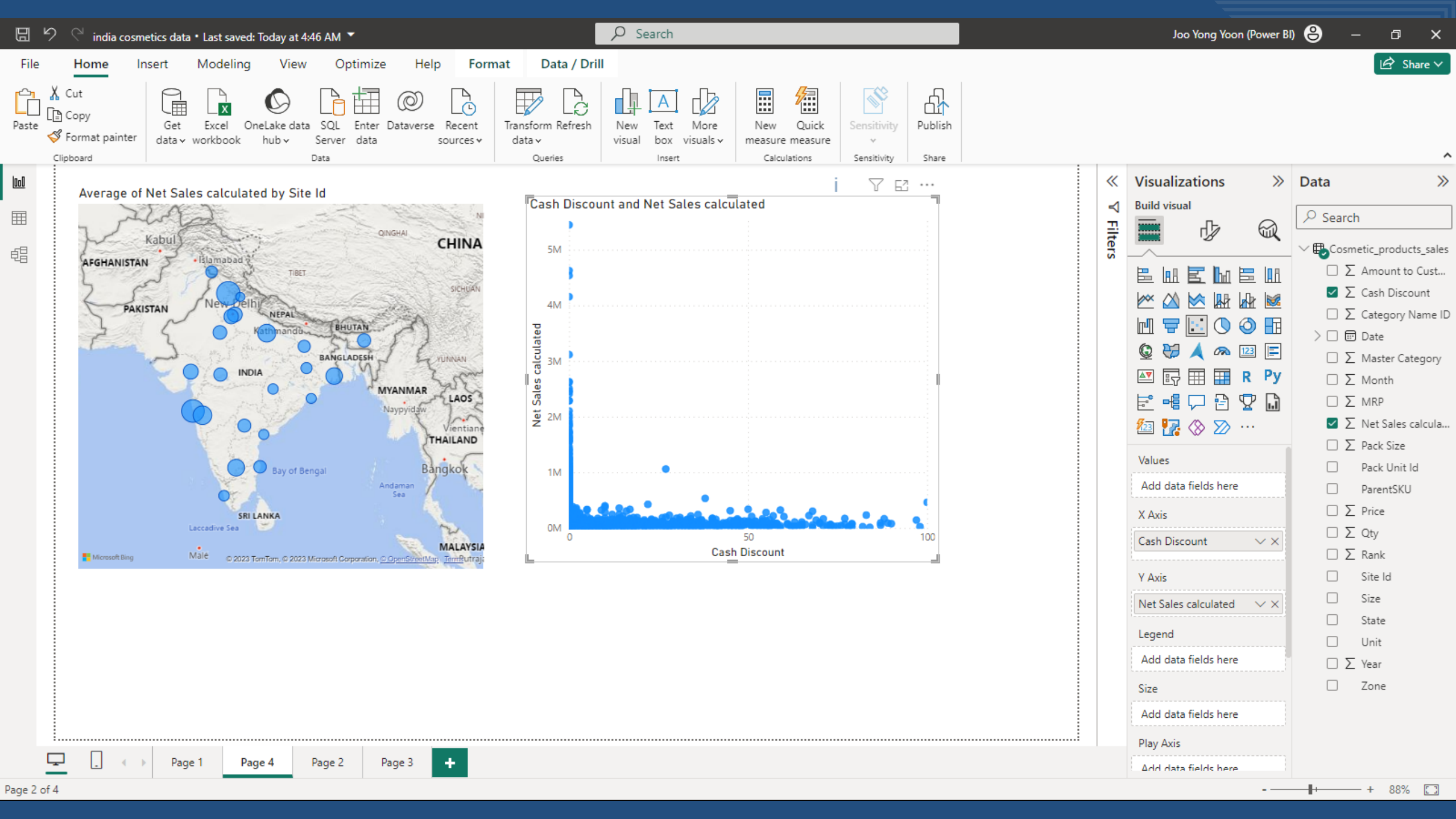
Add data fields here

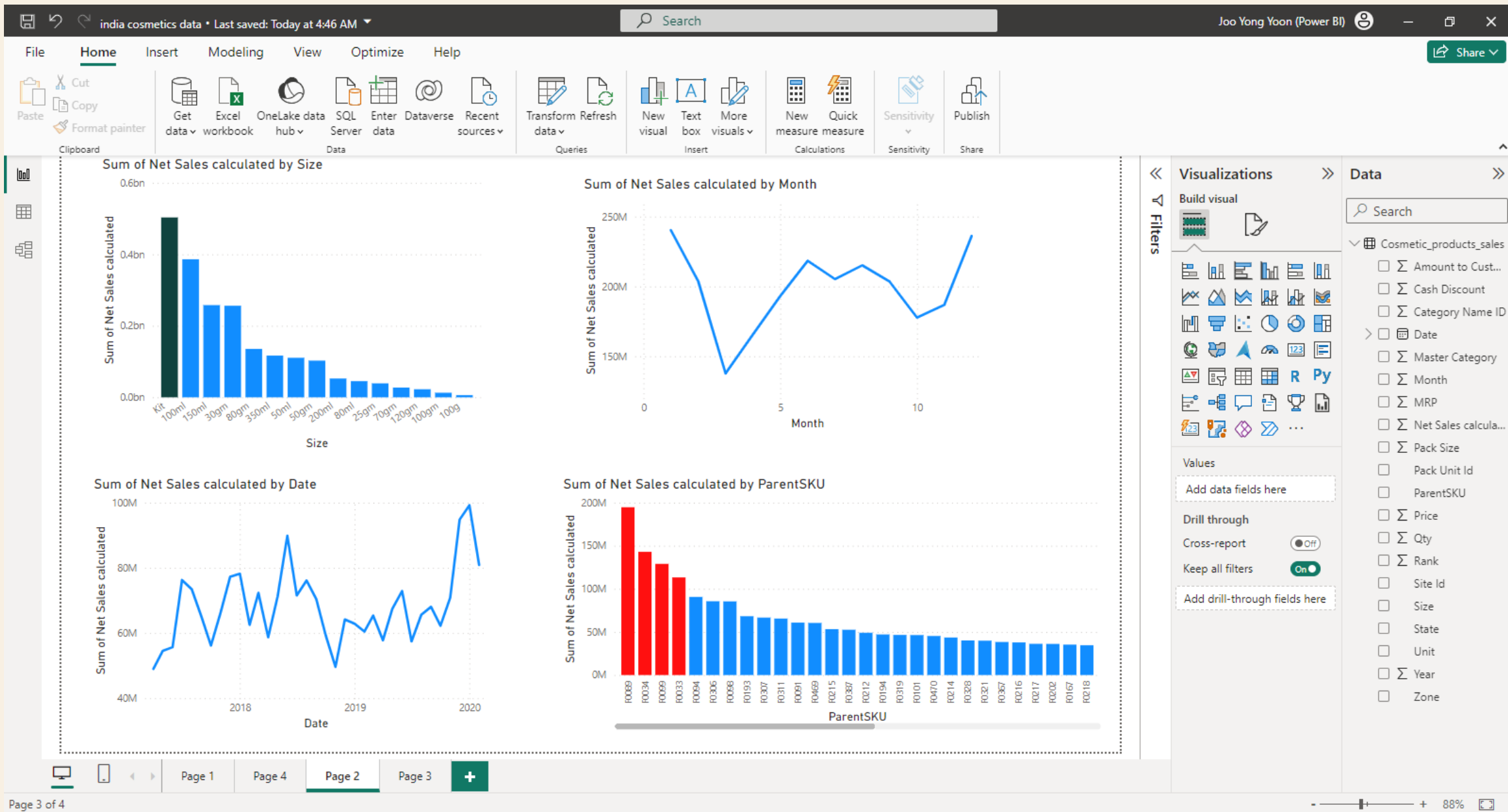
Drill through

Cross-report ☐ Off

Keep all filters ☒ On

Add drill-through fields here





le

HomeInsertModelingViewOptimizeHelp

CutCopyFormat painterClipboard

Get dataExcel workbookOneLake data hubSQL ServerEnter dataDataverseRecent sourcesData

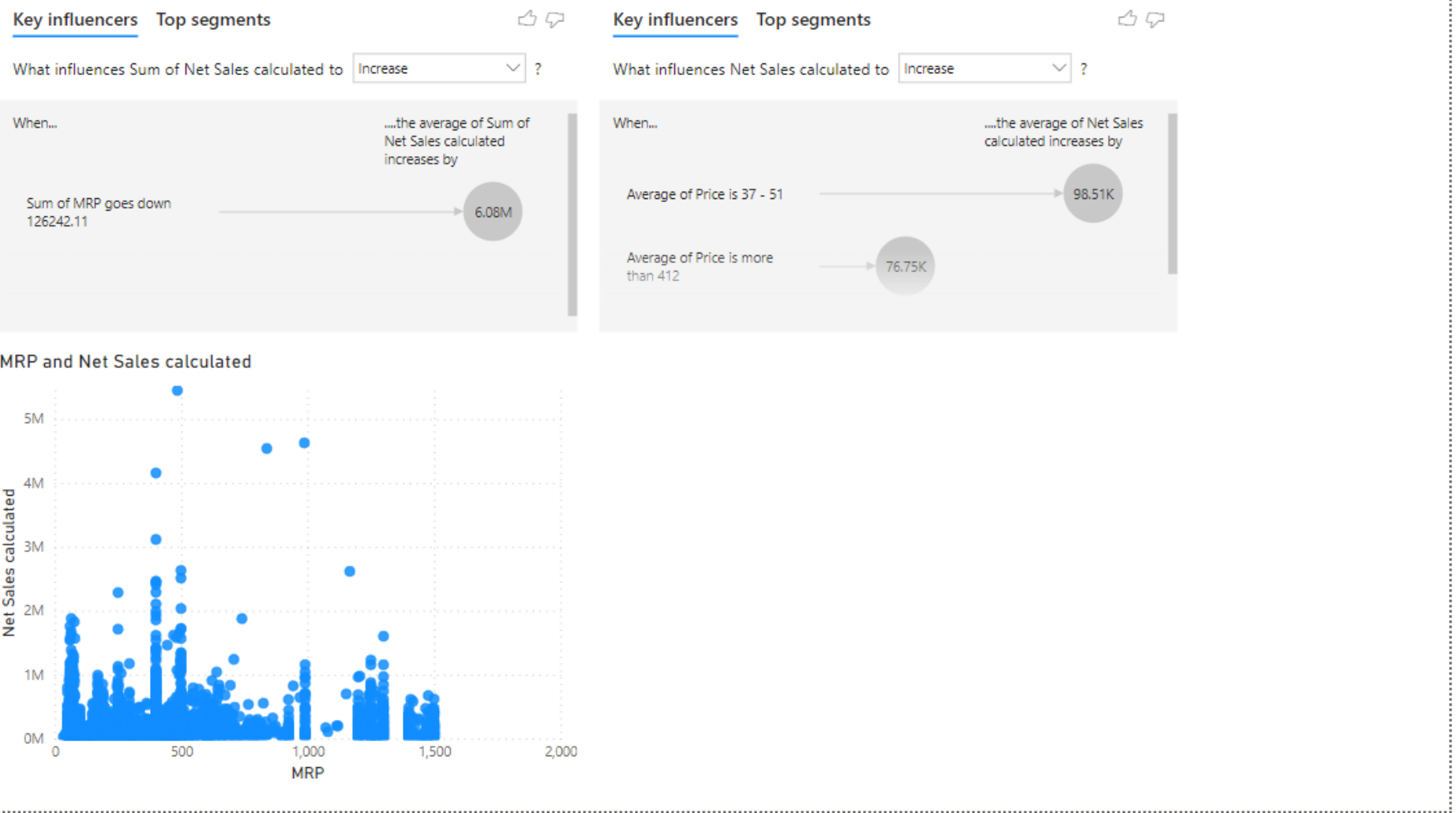
Transform dataRefresh dataQueries

New visualText boxMore visualsInsert

New measureQuick measureCalculations

SensitivitySensitivity

PublishShare



Visualizations

Build visual

Filters

Values
Add data fields here

Drill through
Cross-report (Off)
Keep all filters (On)
Add drill-through fields here

An abstract geometric design on the left side of the slide. It features a dark blue background with various geometric shapes and patterns. A white circle is positioned near the top left. Below it, a light blue semi-circle is visible. To the right of the semi-circle, there is a pink triangle with diagonal lines. Further down, there is a pink square with a pattern of concentric lines. At the bottom, there is a pink triangle with a pattern of concentric lines. The overall design is modern and minimalist.

THANK YOU