COHORT ANALYSIS ON CUSTOMER PURCHASES

Consultant : Deconstruct

Nikesh Mekanaboina

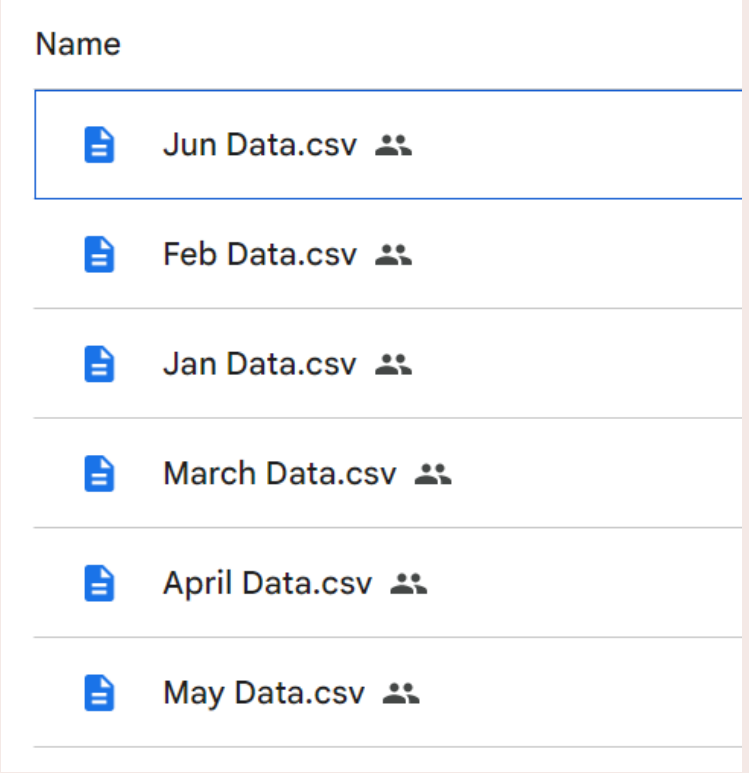
PROBLEM STATEMENT: DATA CONSISTS OF CUSTOMER'S PURCHASES ON AMAZON. IDENTIFY CUSTOMERS TO PERFORM CUSTOMER ENGAGEMENT AND COHORT ANALYSIS

Data Description

- Monthly transaction data from January to June.
- Data consists primarily of customer Email ID, Purchase Date, and Shipping Details.

Key Deliverables

- Total no of orders placed for each month.
- Number of orders from repeat customers.
- Obtain 3-month Cohort percentage
- Present 3-month cohort percentage of each month with a summary.



ACTION PLAN

- Obtain an action plan to structure the solution for obtaining cohort analysis of Amazon purchases data.
- Data Preprocessing
- Data Transformation
- Cohort Analysis
- Obtain Key Deliverables
- Visualize the results
- Summary

Tools used : Jupyter Notebook

Language: Python

DATA PREPROCESSING

- Load monthly datasets into Jupyter Notebook.
- Perform Data cleaning steps to check duplicate and missing values.
- Create a new column month in each dataset.
- Concatenate all datasets into a single dataset to perform further analysis

```
import pandas as pd
Jan_data = pd.read_csv("E:\Desktop\HSLE\ONDC_1\Desc_cohort\Jan Data.csv")
```

```
Jan_data.head()
```

```
March_data = pd.read_csv("E:\Desktop\HSLE\ONDC_1\Desc_cohort\March Data.csv")
```

```
March_data.head()
```

	Amazon Order Id	Merchant Order Id	Shipment ID	Shipment Item Id	Amazon Order Item Id	Merchant Order Item Id	Purchase Date	Payments Date	Shipment Date	Reporting Date
0	408-0462889-4507555	NaN	UgtNLDctH	DCcv5M6BGBL1R	57715685113091	NaN	2023-03-31T18:47:19+05:30	2023-03-31T23:57:16+05:30	2023-03-31T23:57:16+05:30	2023-04-01T02:57:25+05:30
1	408-3886791-6752314	NaN	UC2Q4FH5H	DTJC5W6mG5L3R	34479138059515	NaN	2023-03-30T22:26:11+05:30	2023-03-31T23:53:18+05:30	2023-03-31T23:53:18+05:30	2023-04-01T02:53:25+05:30
2	405-4835411-1629104	NaN	UwT1hykjH	DLn65gBMGXLmR	35122321305339	NaN	2023-03-30T23:41:28+05:30	2023-03-31T23:52:09+05:30	2023-03-31T23:52:09+05:30	2023-04-01T02:52:20+05:30
3	402-2725296-9580366	NaN	UVTLTkckH	D1DKgL6XGZL7R	14511550986291	NaN	2023-03-31T19:57:58+05:30	2023-03-31T23:45:58+05:30	2023-03-31T23:45:58+05:30	2023-04-01T02:46:05+05:30
4	403-4977737-9951524	NaN	UGv496c9H	D8bhgV6xGPLXR	70315686867843	NaN	2023-03-31T14:29:45+05:30	2023-03-31T23:43:29+05:30	2023-03-31T23:43:29+05:30	2023-04-01T02:43:36+05:30

```
#Data Cleaning
```

```
Jan_data.nunique()
```

...

```
Jan_data.isnull().sum()
```

...

```
Feb_data.nunique()
```

...

```
All_data = pd.concat([Jan_data, Feb_data, March_data, April_data, May_data, June_data], ignore_index=True)
```

```
All_data.head()
```

	Amazon Order Id	Merchant Order Id	Shipment ID	Shipment Item Id	Amazon Order Item Id	Merchant Order Item Id	Purchase Date	Payments Date	Shipment Date	Reporting Date
0	407-7394931-1743538	NaN	Uwg1B4knN	DHTCQYHCW	19508303011923	NaN	2023-01-31T14:52:22+05:30	2023-01-31T22:58:44+05:30	2023-01-31T22:58:44+05:30	2023-02-01T01:58:51+05:30
1	407-3968658-3864340	NaN	U3hgBskkN	DZmK2clhb	8495412151755	NaN	2023-01-31T14:50:10+05:30	2023-01-31T22:55:18+05:30	2023-01-31T22:55:18+05:30	2023-02-01T01:55:25+05:30
2	402-2499146-7573940	NaN	UCXGy9kHN	DnQBnWzCR	30423700620379	NaN	2023-01-31T15:19:20+05:30	2023-01-31T22:52:17+05:30	2023-01-31T22:52:17+05:30	2023-02-01T02:02:30+05:30

DATA TRANSFORMATION

- Perform feature engineering by extracting the month from the purchase date column.
- Create a cohort month column on the customer who made the first purchase of the month.
- Cohort is classified as it commensurate with the behavior of customers purchase patterns.

```
# Create a cohort based on the first purchase month
All_data['CohortMonth'] = All_data.groupby('Buyer Email')['Purchase Date'].transform('min').dt.to_period('M')
```

```
All_data.head(15)
```

Date	Shipment Date	Reporting Date	...	Item Promo Discount	Shipment Promo Discount	Carrier	Tracking Number	Estimated Arrival Date	FC	Fulfillment Channel	Sales Channel	Month	CohortMonth
-01-5:30	2023-01-31T22:58:44+05:30	2023-02-01T01:58:51+05:30	...	0.0	-33.9	ATS	322236665304	2023-02-03T20:00:00+05:30	PNQ2	AFN	Amazon.in	2023-01	2023-01
-01-5:30	2023-01-31T22:55:18+05:30	2023-02-01T01:55:25+05:30	...	0.0	-33.9	ATS	322236725589	2023-02-02T20:00:00+05:30	BOM7	AFN	Amazon.in	2023-01	2023-01
-01-5:30	2023-01-31T22:52:17+05:30	2023-02-01T02:02:30+05:30	...	0.0	0.0	ATS	157819312265	2023-02-01T08:00:00+05:30	HYD8	AFN	Amazon.in	2023-01	2023-01
-01-5:30	2023-01-31T22:30:24+05:30	2023-02-01T01:41:13+05:30	...	0.0	-33.9	ATS	322236557425	2023-02-01T20:00:00+05:30	MAA4	AFN	Amazon.in	2023-01	2023-01
-01-5:30	2023-01-31T21:10:11+05:30	2023-02-01T00:10:17+05:30	...	0.0	0.0	ATS	322235689974	2023-02-02T20:00:00+05:30	DEX6	AFN	Amazon.in	2023-01	2023-01

COHORT ANALYSIS

- Group customers based on their cohort month where first purchase month with current purchase month.
- Calculate the total no of orders from each month from customers.
- Cohort period is the measure of time since the cohort was formed.
- Cohort period is obtained with the measure of cohort month.
- In summary, cohort analysis helps you understand how different groups of customers evolve over a period, starting from their first interaction with your business.

```
cohorts.head()
```

	CohortMonth	Month	TotalOrders
0	2022-12	2022-12	53
1	2022-12	2023-01	3
2	2022-12	2023-02	3
3	2022-12	2023-03	4
4	2022-12	2023-04	4

```
cohorts.head()
```

	CohortMonth	Month	TotalOrders	CohortPeriod
0	2022-12	2022-12	53	0
1	2022-12	2023-01	3	1
2	2022-12	2023-02	3	2
3	2022-12	2023-03	4	3
4	2022-12	2023-04	4	4

KEY DELIVERABLES

- Total no of orders for each month was obtained.
- Number of orders from repeat customers resulted.
- 3- Month cohort percentage for each month.
- New customers where the absence of purchase history

	Amazon Order Id	Merchant Order Id	Shipment ID	Shipment Item Id	\
21707	405-6272841-2028331	NaN	U6tfN4zJ9	DqcC38SydKLMR	
13940	406-5822595-0963550	NaN	U3GxLTwW4	Dx3ZLpr8R	
17257	171-8785773-8148359	NaN	UgFbnGDkH	D5hGS8T7R	
11582	404-4675362-6776330	NaN	U6PL17hyb	DGhq587Lg	
22870	404-8140931-9073929	NaN	U3vdKRrn9	D6RKqCjfdSLNR	
...	...	...	...	...	
25536	404-4762880-4576340	NaN	UwY7rRp8w	DHn5Pw0NSfLSR	
19440	404-4131717-4397139	NaN	UnkhjBFx9	DLcGxVlW65LPR	
3002	406-4823486-3794752	NaN	UJXPLkBdR	MYTZNDkpj	
7089	406-1838017-0601160	NaN	UVsnpF4bN	X5DGBgdmB	
28116	405-7894812-9040359	NaN	UVZnhZ23B	DTllSbNj6nLQR	

Total No Orders

Outcome 1: Total Number of Orders Placed for Each Month

	Month	Total Orders
0	2022-12	53
1	2023-01	3205
2	2023-02	3629
3	2023-03	5802
4	2023-04	5327
5	2023-05	5578
6	2023-06	6151

Repeat Customer orders

Number of Orders from Previous Customers in the Coming Months for Each Month

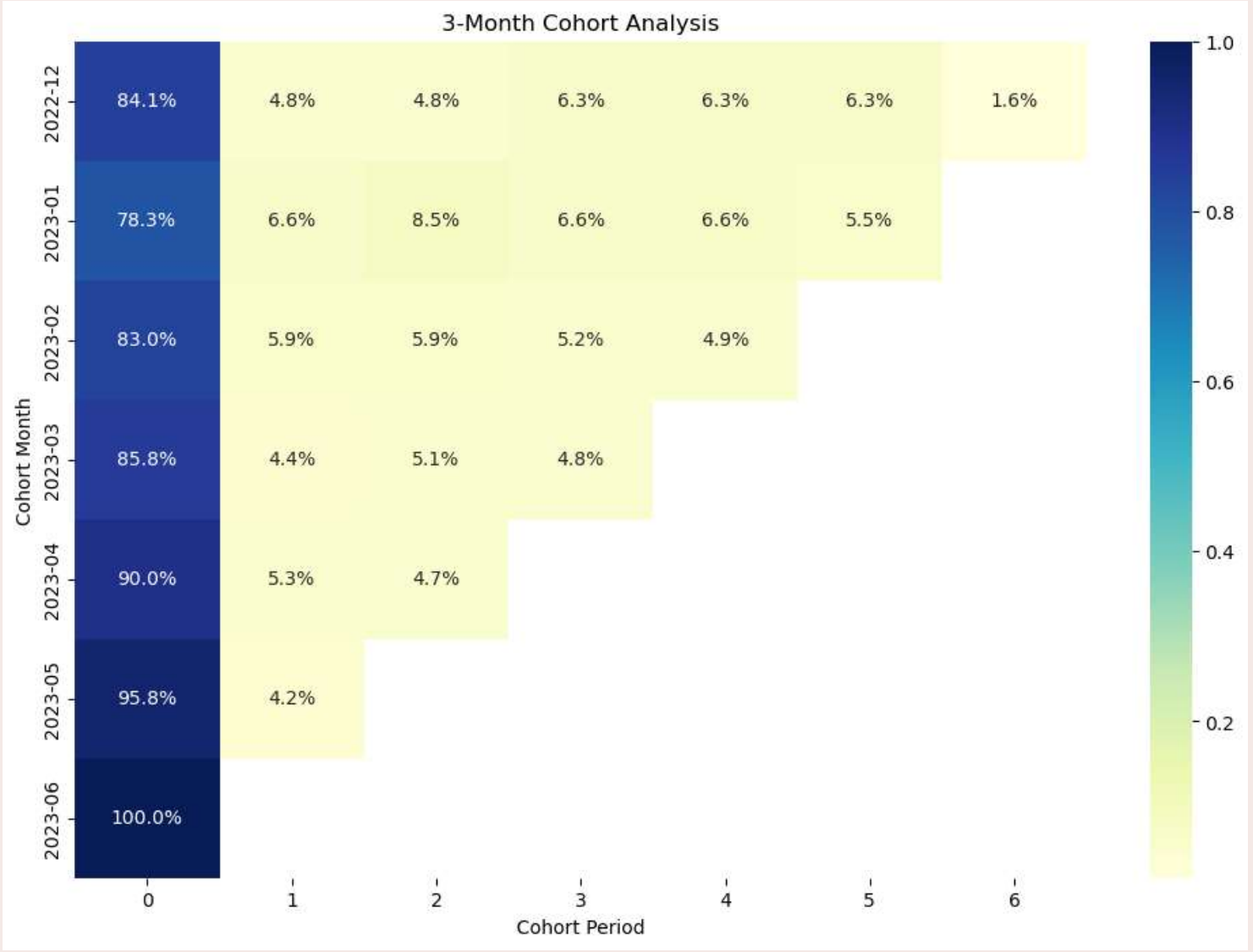
	Month	Orders
0	2022-12	53
1	2023-01	3036
2	2023-02	3432
3	2023-03	5422
4	2023-04	5050
5	2023-05	5264
6	2023-06	5857

3- Month Cohort Percentages

3-Month Cohort Percentages:

	CohortMonth	3MonthCohortPercentage
0	2022-12	0.875000
0	2023-01	0.891774
0	2023-02	0.953582
0	2023-03	1.000000
0	2023-04	1.000000
0	2023-05	1.000000
0	2023-06	1.000000

3 MONTH COHORT ANALYSIS CHART



SUMMARY

- First 3 month cohort of Total no orders were obtained as average orders from customers and the preceding 3- month cohort had majority of orders.
- To tailor market segmentation and further analysis repeat customers made purchases more on the preceding 3-month cohort(April-June).
- We observed incremental growth in repeat customers in January – 89% and February – 95%.
- Additionally, there is an increase of new customers exponentially from month wise.
- 3-month cohort percentage serves as a vital indicator of customer loyalty and repeat business.
- The actionable insights derived from this analysis pave the way for informed strategies aimed at nurturing customer relationships and enhancing overall business performance.