Cohort based Monthly Retention using 2 datasets Retail and Restaurants

- A food delivery platform that makes it easy for hungry people to order great food from Restaurants in addition it has expanded the offering beyond food into Retail: ranging from grocery to clothes, flowers, beauty, and other sectors!

Presented by:

Gracy Jason Joseph

Dialect and tools used:

1. **Query:**
   PostgreSQL with pgAdmin

2. **Additional tools:**
   Jupyter Notebok

3. **Visualization:**
   Chart type: Heatmap to show user retention

Cohort based monthly retention for Retail product line

Monthly cohort of users who made their first and possibly following purchases for Retail:

Cohort_month: Grouping the users based on month

Total_users: 43 users made the first purchase in March

Num_users: 7 users made purchase in the next month April after their 'first purchase month'

Month_number: the month after the first_purchase month eg. month_number 1 represents April

Percentage: percentile of number of users (7/43)

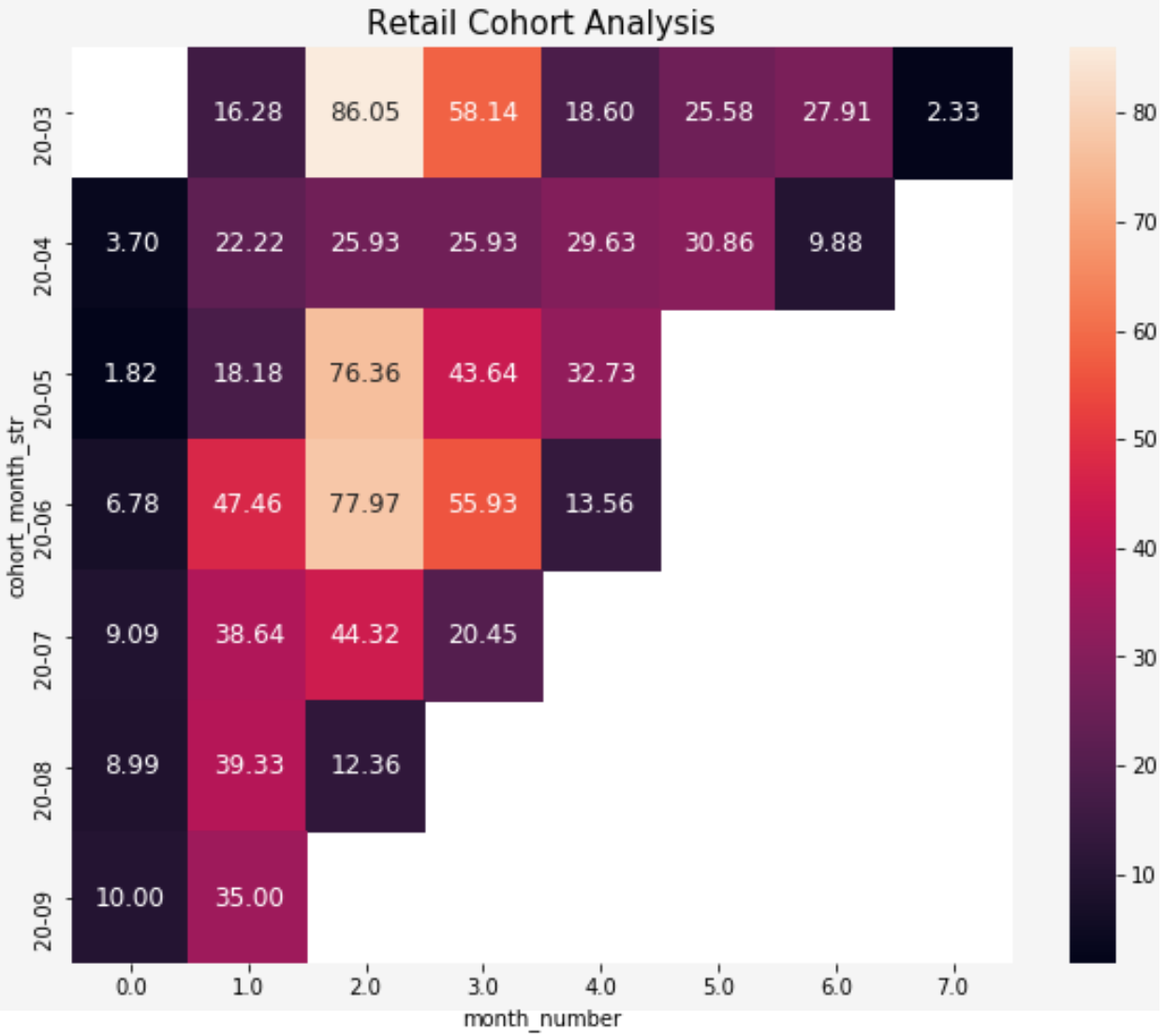
	cohort_month	total_users	num_users	month_number	percentage	cohort_month_str
0	2020-03-01	43	7	1.0	16.28	20-03
1	2020-03-01	43	37	2.0	86.05	20-03
2	2020-03-01	43	25	3.0	58.14	20-03

Cohort based monthly retention for Retail product line

Monthly cohort of user engagement who made their first and possible following purchases for Retail:

- From the retention chart, 7 users made the purchase on April 2020 after their first purchase month March and the retention was 16.28%, and on month 6 it was 27.91% and by the 7th month retention was 2.33%. So on the 6th month the users who made their purchase after their first month were still active users/customers on the Wolt.

- Out of all the new users/customers during the time range, there are active users who are retained from their first purchase month to the 7th cohort month.



Cohort based monthly retention for Restaurant product line

Monthly cohort of users who made their first and possibly following purchases for Restaurant:

Cohort_month: Grouping the users based on month

Total_users: 8157 users made the first purchase in March

Num_users: 1588 users made purchase in the next month 'April' after their 'first purchase month'

Month_number: the month after the first_purchase month eg. month_number 1 represents April

Percentage: percentile of number of users (1588/8157)

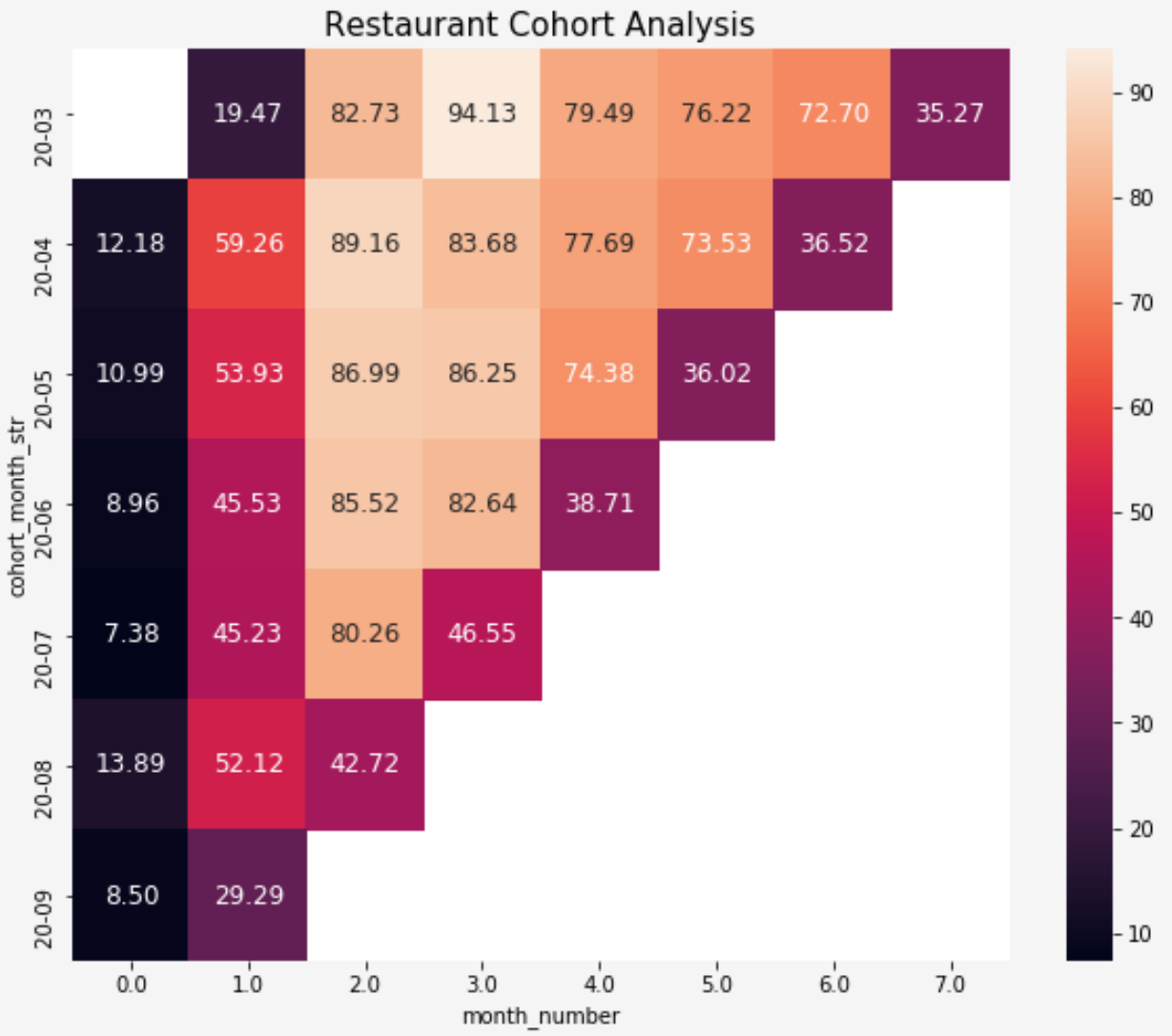
	cohort_month	total_users	num_users	month_number	percentage	cohort_month_str
0	2020-03-01	8157	1588	1.0	19.47	20-03
1	2020-03-01	8157	6748	2.0	82.73	20-03
2	2020-03-01	8157	7678	3.0	94.13	20-03
3	2020-03-01	8157	6484	4.0	79.49	20-03

Cohort based monthly retention for Restaurant product line

Monthly cohort of user engagement who made their first and possible following purchases for Restaurant:

- From the retention chart, 1588 users made the purchase on April 2020 after their first purchase month March and the retention was 19.47%, and on month 6 it was 72.70% and by the 7th month retention was 35.27%. So on the 6th month the users who made their purchase after their first month were still active users/customers on the Wolt same as of Retail product line.

- Out of all the new users/customers during the time range, there are active users who are retained from their first purchase month to the 7th cohort month.



What assumptions about the data did you make when creating your retention charts?

1. Lack of customer engagement shouldn't be hidden/masked by the impressive growth of new customers as there should be customer/user engagement, loyalty and continuation of services over the long run

2. **Product Lifetime (depicted vertically in the chart):**

   Percent of customers who are loyal and engaging after the 1st cohort month. The lifetime months can be due to the quality, experience, performance of the Wolt

3. **User lifetime (depicted vertically in the chart):**

   The long term relationship of the customers in any cohort month. The chart shows how long the customers are continuing the service. This can be due to the quality of products, operations and customer/team support.

4. By creating and comparing the retention chart, we can track the customer behavior over a period of time or how the same behavior differs for different cohort month

Did you encounter any problems in the data?

While working on the data I found that ‘User id’ in purchases dataset didn’t have first purchase date.

```
select distinct p.user_id, purchase_date,
first_purchase_date
from purchases p left join first_purchases fp
on p.user_id=fp.user_id
```

Output | Explain | Messages | Notifications

user_id integer	purchase_date date	first_purchase_date date
7161737	2020-07-28	2020-03-16
8420792	2020-09-05	2020-04-25
9802117	2020-10-25	2020-05-13
5927352	2020-07-22	2020-03-21
1339244	2020-09-18	2020-05-20
4802056	2020-08-02	2020-05-02
8070918	2020-08-21	2020-07-08
4314849	2020-10-21	2020-09-02
7891798	2020-05-28	[null]
1206632	2020-07-26	2020-06-10
3643694	2020-07-26	[null]
8859640	2020-06-26	2020-03-26

Which additional data sources or business information do you think could be used to improve your solution?

1. Understanding customer retention wrt. each product line

2. **Drawing conclusions based on product category:**

1) Customers ordering food from 'Restaurant' product line in the first purchase month showed higher retention than the rest.

- Some possible inferences can be that the quality, experience, marketing aspect for the restaurants needs to be consistent.

2) Customers buying clothes from 'Retail' product line in the first purchase showed the lowest retention rate.

- The retention strategies for clothes purchasers need to be relooked.

3) 1st month is critical or dropping as the customer churn seems to increase beyond that.

- Retention strategy for users entering the next months since their first purchase can be evaluated.

Additional approach to retain customers month after month

1. By conducting Customer satisfaction surveys to help know the customer metrics according to the product line.

2. Finding the problem source: lack of customer data, making note of customer feedback forms, lack of information on product, making it easy for the customers to report an issue faced in either of the product line.

3. Analyzing customer activities over the period of time mainly cancellations, inactivity, changes in plans.

4. Evaluating products, services and also expectations provided to the customers as it shouldn't be over promising. or under delivering

Thank You!