Assessment

1. The marketing team wants to understand our vendors a bit better. Their questions are:

a) Is there any order imbalance per vendor in each market?

b) Could we identify some relevant segments of vendors?

c) Are there any other interesting insights you would share with them?

Use the orders data provided. Given the request and the available data:

1. With a tool of your choice (R/Python, SQL etc), use the shared data to build an analysis that could answer their questions.

2. Aside from the available data, what other features would you include for an effective segmentation?

3. How could the marketing team use this output to improve their marketing performance?

a) There has been slight imbalance has been observed in some of the cases

```
In [4]:  import pandas as pd

In [12]: df1 = pd.read_csv('Recruitment_Assessment/dim_order_details.csv')
         df2 = pd.read_csv('Recruitment_Assessment/fct_orders.csv')

In [18]: output1 = pd.merge(df1, df2, on='orderid', how='inner')

In [21]: print(output1)
```

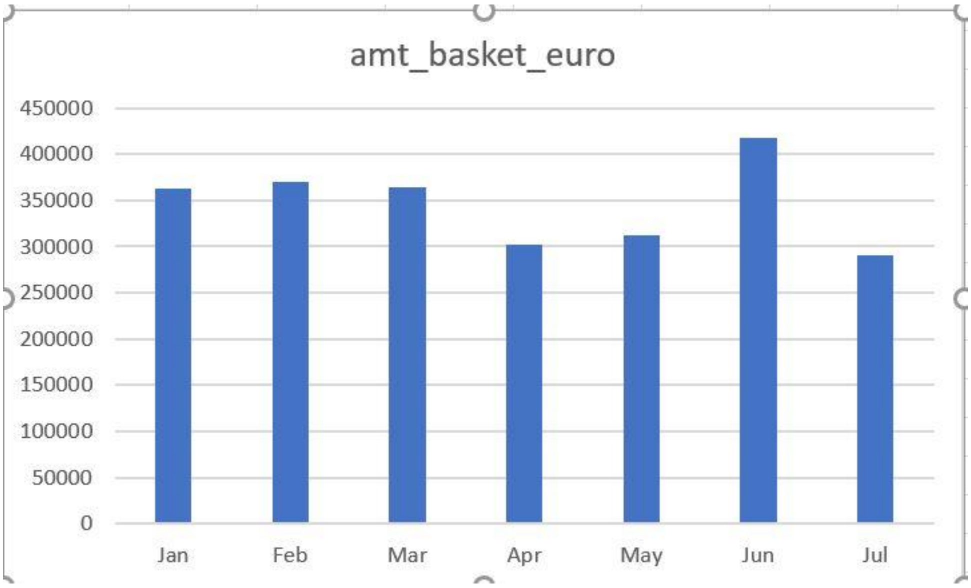
```
         orderid                  order_date  is_voucher  is_discount  \
0          61411  2020-01-01 01:05:24.000000       False        False
1          65675  2020-01-02 17:18:30.000000       False        False
2          52408  2020-01-04 17:59:42.000000       False        False
3          98735  2020-01-06 09:52:24.000000       False        False
4          52313  2020-01-07 19:49:35.000000       False        False
5         111115  2020-01-09 17:15:16.000000       False        False
6          48054  2020-01-11 15:11:08.000000       False        False
7          48060  2020-01-12 22:13:50.000000       False         True
8          98764  2020-01-14 19:46:49.000000       False        False
9          56717  2020-01-16 17:40:36.000000       False        False
10         65777  2020-01-18 14:35:51.000000       False        False
11         52507  2020-01-20 14:25:12.000000       False        False
12         36249  2020-01-22 14:49:56.000000       False        False
13         78169  2020-01-24 19:53:32.000000       False        False
14         36401  2020-01-26 00:10:21.000000       False        False
15         65716  2020-01-27 19:22:04.000000       False        False
16        111035  2020-01-28 20:55:53.000000       False        False
17         36278  2020-01-30 14:20:59.000000       False        False
```

```
In [22]: output1.duplicated(subset=['customerid'])

Out[22]: 0     False
         1     False
         2     False
         3     False
         4     False
```

b) The relevant segments of vendors are the vouchers that has been used by first time users which can help in getting better conversion rate.
The other is the area of discount segment which are given to the active users. The other area which needs to be taken into account for the churning users before they lapse by giving them promotions, discounts, new features of application.

c) The graph below shows monthly transactions which shows the highest for the month of June and lowest for the month of July. The marketing team can use this information to keep a track on first time users, active users and churning users. The detailed answers has been given below.



Answer

2. For achieving better conversion and retention rates, an effective segmentation strategy is essential which can be done by identifying the features of users and then grouping users accordingly.

The main segments for delivery apps are First-Time Users, Active Users, Churning Users

Segment: First-Time Buyers

For the users who have installed the application but haven't made any purchase until. The intention is to activate those users by giving incentives/discounts/content(e.g. products or restaurants) in order to make them purchase using the application.

Segment: Active Buyers

For the users who have ordered in past 30 days which qualifies them as customers who are active or somewhat interested. The intention is to expand the frequency of orders per user. Let's suppose if we have customers who place orders twice a week, the goal is to increase the frequency of orders to at least three to four times or more so we can turn active buyers into loyal users.

Segment: Churning Buyers

For the users who have not been really active; their last order was 30 days ago. That means these users are slowly loosing interest or fading away and if not retargeted in a certain way might uninstall the application. The intention is to bring these churning users by giving them some new incentives to make them order one more time

3. The marketing team can use this output to improve their marketing performance for each user segment by setting targets.

For budget allocation to convert users depends on where they are in the funnel and the price must pay accordingly. As all the customers are all on the different stage of customer lifestyle, the targets must be set accordingly for all the stages. The aim is not to prioritize one segment over the other but to find a solution in a most cost-effective way possible by distributing the investment segment-wise.

- First-Time Buyers: In comparison to other segments, the investment is likely 10 times higher than the other segments as this the most crucial stage where the interest of the customer is at the highest. The first-time order ultimately finalizes the user acquisition journey which is called lifetime value (LTV) of the buyers to determine the target. For example, if it's a growth stage and buyers' average LTV in the first year is $300, this is a target goal which needs to be achieved.
- Active Buyers: To drive loyalty from the buyers, the target for this segment is to maximize break-even point. It depends on how much investment can be done for an additional order placed by an active buyer. For example, the application has a margin of $2 per order, the target on the network

advertisements would be $2 to get that order. Each of the order placed by the users, taking the price per order into account, if 10 orders triggered at $2 then $20 would be the investment for this segment.

- Churning Buyers: If the immediate action has not been taken then the one is likely to lose the users in the segment and one needs to reinvest in reactivating the application one more time. For this segment, the price point is between the price one is willing to invest for an active buyer and for an inactive buyer. It is bit harder to bring back the lapsing users in comparison to the ones in other segments. For this campaign to make more financial sense, the wise decision would cost less for lapsing users than acquiring new users.

  The first order satisfaction is the key to retaining buyers. This means the ads shown to the first-time buyers not only focus on first-time conversion but also for buyer's retention. As the first-time placed order is only the beginning of the customer retention journey so it is important to have a quality experience of the application. By giving the previous purchasers, as journey-based contents.

  For First-Time Buyers, showing static banners for top vendors, associated promotions.
  For Active Buyers, displaying dynamic product ad with vendors or orders based on previously made purchase.
  For Churning Users, displaying products ads primarily focusing on top vendors with deals, static ads with upcoming restaurants or new verticals launched, or highlighting new features with the application.

  This approach is adaptable for the active and lapsing users. By testing and measuring such a setup will help in optimizing each segmentation tactics towards performance and scale. Focusing on the customer's lifecycle, this strategy helps in reaching to each and every customer in the database by providing them with relevant creatives to drive conversion rate and long-term loyal

2) One of our analyses revealed that 30% of the new users ordering for the first time from popular brands (such as McDonalds/BurgerKing) never placed a second order on our platform.

At a first glance, this seems rather high and we want to do a second analysis to better understand these users. To give a direction for the new analysis, we would need to start by compiling a list of assumptions about these users.

a) Could you think about 1-2 assumptions you would like to focus on?

b) What kind of data would you look at in order to test these assumptions?

Answer

a) Being a user Talabat, I would say the possible assumptions that why people didn't place the order second time cuz popular brands have higher delivery fees in comparitively to other restaurants
b) Distance between the branch of a restaurant and residence of th customer

The likely options are customers would either do take aways by placing an order directly in a branch or do drive throughs

The data we would like to look at in order to test the assumptions:

Delivery fee of popular restaurants by using country code and chain_id

Distance of the residence of the customer and the branch by using features like longitude and latitude

3) **Your colleague ran a linear regression analysis to understand the effect of talabat vouchers on the 1st visit on the total number of sessions (the training data was consolidated on a user level, with voucher = binary variable, with values 0/1 and target variable = total number of orders during the entire lifetime of each user). The results showed an R2 = 0.6 and the estimated coefficient for the voucher was 5.**

He needs help interpreting these results and checks in with you before presenting the results to the shareholders. What would you advise?

Answer

The tight set of data will have a regression line that's close to the points and have a high level of fit, meaning that the distance between the line and the data is small. Although a good fit has an R2 close to 1.0, this number alone cannot determine whether the data points or predictions are biased.

The usefulness of R2 is its ability to find the likelihood of future events falling within the predicted outcomes. The idea is that if more samples are added, the coefficient would show the probability of a new point falling on the line.

Even if there is a strong connection between the two variables, determination does not prove causality.

The higher the coefficient, the higher percentage of points the line passes through when the data points and line are plotted. If the coefficient is 0.80, then 80% of the points should fall within the regression line. Values of 1 or 0 would indicate the regression line represents all or none of the data, respectively. A higher coefficient is an indicator of a better goodness of fit for the observations.

Since the estimated R2 is 5 and our model is showing R2 0.6 which shows that our model is goodness of fit model and has done better than expected.

4) **The management team wants to accelerate conversion of customers who placed an order only on food to new verticals recently launched: grocery & tmart (talabat-owned grocery shops).**

a) What approach would you suggest considering that there is a small portion of customers who have converted already from food to new verticals?

b) Please suggest a way you could provide a set of insights to our CRM team for setting up new experiments that can validate your approach.

Answer

a) The research has been done based on the reviews of Talabat grocery application and also I have done a little verbal survey asking my friends around on how their experience is with grocery vertical on Talabat. The common problems which have been seen mostly are the following:

- Missing items

- Quality of the products

- Packaging

And the positive feedback includes fast delivery.

b) Unlike the current model where Talabat's groceries are linked to multiple super-markets, the new experiment could be done by limiting it to suppliers plus having Talabat's **own range of groceries** as this could:
   - Ensure the quality

- Missing items problems will be addressed more accurately by the customer support. Instead of refund or replacing the items, delivery will only be done once all the items are ready
- Providing customers with non-biodegradable packaging
- Good everyday deals

5) One of our brands, OTLOB, was recently merged to be part of talabat. This required a migration for part of the OTLOB existing customer base onto the new platform. Management would like to get a view on how customer behaviour evolved after the migration.

a) What dimensions & metrics would you use to monitor the pre/post migration analysis?

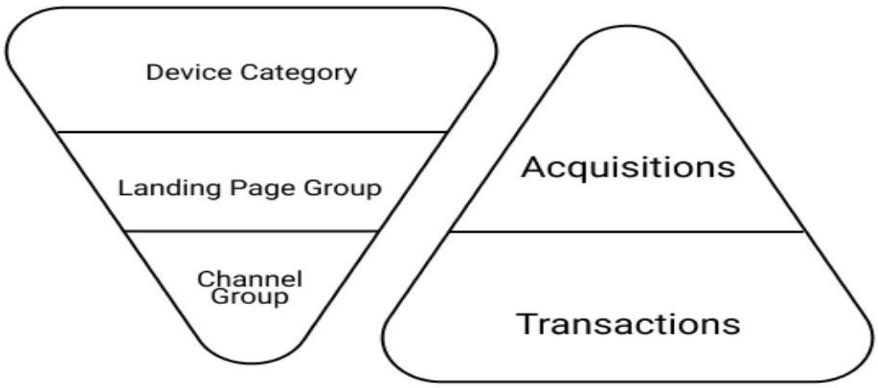
b) How would you present the insights results to management? Feel free to create dummy data to demonstrate how you would tell the story.

Answer

a) Dimensions and Metrics

   The widely used metrics by most of the marketing professionals is the Conversion Rate metrics as it is one of the metrics which includes both input and output data. A conversion of 40% is generally considered as high-performance metrics.

The diagram below shows dimension and metrics to start with conversion analysis



The above set of dimensions and metrics is useful for the marketing team as well as product team.

In **Metrics, transactions** are the ultimate result for every session.

Acquisitions are the first order placed by the new user. It is one of the mot valuable metrics for the growth of the business. The platform that can easily acquire new users affordably and can retain them wins the game. Acquisitions are way less than the transactions, so their segment is not as deep as transactions.

Talking about **Dimensions**, there are 3 major dimensions such as **Device Category**, **Landing Page**, **Channel Group.**

The performance of the **Device** mostly differs as it gets separated into big buckets such as website, mobile website plus tablets, IOS & Android apps. Hence, the performance of each varies because:

- IOS & Android application requires installation, so App store or Google Play Store is a part of conversion flow
- Website targets users that do their transactions from laptops/tablets/PC's while mobile website targets users that use their phone and does not prefer the apps.

Landing Page grouping stands for landing page groups. Every group of landing page has different behavior. For example,

- Group of restaurants in a specific location (e.g., Dubai)
- Group of restaurant part of big brand chain (e.g., Mc. Donald's, KFC, Hardee's)
- Group of specific cuisine restaurants (e.g., Pizza)

Lastly, **Channel Grouping** is the way of grouping Marketing channels. The main difference between channels is that there are channels that have intention of high purchases and there are channels whose traffic is known for brand performance with a relatively high conversion rate.

These dimensions could be used in combination or as stand alone. The conversion can be checked by each marketing channel or conversion rate of users that land on your mobile website's homepage via direct source and dig deeper into measuring device categories performance.

b) Since conversion rate is best friend, we overview performance using conversion rate along with transactions or revenue as first step.

In the picture above, the performance seems very healthy. Though the conversion rate is seeming steady, sessions and order increase linearly. By investing the right marketing spend and product optimization, this can be scaled easily.

The daily report of the seasonality factor by taking in consideration revenue example. Regarding transactions Thursday, Tuesday, Wednesday are the best days while Sunday and Saturday are the worst. But regarding conversion rates Monday, Sunday and Wednesday are the best days. Wednesday and Sunday have a 0.12% conversion rate difference but Wednesday has almost double revenue

Traffic per day in homepage

Channel	Sessions	Transactions	E-commerce conversion rate	Revenue
	59,662	1,743	2.92%	336,361.47
Thursday	10,960	290	2.65%	55,489.50
Tuesday	10,573	310	2.93%	54,263.42
Wednesday	10,446	341	3.26%	59,970.54
Friday	8,855	252	2.85%	56,795.96
Monday	7,978	254	3.18%	53,389.90
Sunday	5,486	172	3.14%	34,992.06
Saturday	5,364	124	2.31%	21,460.09

6)

1. How many unique users visited the website that day?

```
#query1
#How many unique users visited the website that day?
SELECT count(distinct fullvisitorid) as Unique_Users From `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`
```

Row	Unique_Users
1	2293

2. How many sessions did users start on average?

```
#query2
#How many sessions did users start on average?
SELECT count(distinct concat(fullvisitorid, cast(visitstarttime as string))) / count(distinct fullvisitorid) as
number_of_sessions_per_user, FROM `bigquery-public-data.google_analytics_sample.ga_sessions_20170801` LIMIT 1000
```

Row	number_of_sessions_per_user
1	1.114696903619712

3. What is the website's conversion rate? (To calculate conversion rate do not consider more than one transaction per session)

```
#query3
# What is the website's conversion rate?

Select count(distinct hits.transaction.transactionid) / count(distinct concat(cast(fullvisitorid as string), cast(visitstarttime as
string))) as ecommerce_conversion_rate From `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`, Unnest(hits) as hits
```

Row	ecommerce_conversion_rate
1	0.016823161189358372

4. What is the conversion rate per traffic medium?

```
#query4
#What is the conversion rate per traffic medium?
select hits.page.pagepath as goal_completion_location, count(distinct(if(regexp_contains(goal_completion_location,
'/ordercompleted'),session_id,null)))/ count(distinct concat(cast(fullvisitorid as string), cast(visitstarttime as string))) as
traffic_medium_conversion,
  count(distinct(if(regexp_contains(goal_completion_location,'/basket.html'),session_id,null)))/ count(distinct concat(cast(fullvisitorid as
string), cast(visitstarttime as string))) as traffic_medium2
  from `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`,
    unnest(hits) as hits
```

(Some issues faced running this query)

5. How many minutes on average did it take the users to reach the checkout confirmation page (title = 'Checkout Confirmation')?

-

```
#query5
#How many minutes on average did it take the users to reach the checkout confirmation page (title = 'Checkout Confirmation')?
SELECT sum(h.minute) /count( h.ecommerceaction.action_type = 'Checkout Confirmation')as product_adds_to_cart,  FROM `bigquery-public-
data.google_analytics_sample.ga_sessions_20170801`, UNNEST(hits) as h
```

Row	product_adds_to_cart
1	29.305750774578705

6. The website tracks add and remove from cart interactions using two events (Action = 'Add to Cart' and Action = 'Remove from Cart'). Please use that information and investigate the data further to answer the following questions:

a. How many products in total did users add to cart?

```
#query6(a)
#How many products in total did users add to cart?


SELECT count(h.eCommerceAction.action_type = 'Add to Cart' ) as product_checkouts, FROM `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`, UNNEST(hits) as h
```

Row	product_checkouts
1	13233

b. In sessions where users actually added products to their carts, how many products were added per session on average?

```
# query6 (b)
#In sessions where users actually added products to their carts, how many products were added per session on average?
SELECT sum(productquantity) / count(h.ecommerceaction.action_type = 'Add to Cart') as product_adds_to_cart FROM `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`, UNNEST(hits) as h, unnest(product) as product
```

Row	product_adds_to_cart
1	0.10219391069295727

c. What is the most added-to-cart product and what is the most removed-from-cart product (mention the product name)?

```
#query6c
#What is the most added-to-cart product and what is the most removed-from-cart product (mention the product name)?


SELECT count( h.ecommerceaction.action_type = 'Most Added To the Cart')product_adds_to_cart, count(h.ecommerceaction.action_type = 'Remove From The Cart') as product_removes_to_cart, v2ProductName as Product_Name FROM `bigquery-public-data.google_analytics_sample.ga_sessions_20170801`, UNNEST(hits) as h, unnest(product) as product   group by v2ProductName
```

Row	product_adds_to_cart	product_removes_to_cart	Product_Name
1	104	104	Electronics Accessory Pouch
2	346	346	Google Flashlight

Bonus Questions

1. Did you find any insights / observations while you were investigating the data that you think are interesting to share?

2. Do you have any thoughts, notes, doubts or suggestions on the data structure?

The insights are the fields having null and missing values which might not show the true picture when it comes to doing quantitative and qualitative analysis.