

Table of Contents

Case Study 1: Expanding a Crypto Bank Offering a Global Card	3
Approach, choices, thought process	3
Dashboard	4
Exploratory data analysis, Interpretation of results and insights	5
General Insights	5
Transactions	9
All Tables (joining up the data)	10
a. Transactions	10
Transactions Mcc's	11
Top 1% Transactions	12
First transactions	13
Cohort Analysis	14
b. Users	15
c. Cards	16
Segments	18
1) RFM	18
Hypothesis testing	20
2) Spending Segments ("Whales" Method):	21
Markets	22
1) Entering new markets (we don't have users there):	22
2) Expanding current markets where we have good users:	25
High spending segment (Countries with most of this users)	26
RFM	27
Credit cards	28
Biggest spending	29
Highest ratios of users with cards and transactions (%):	29
Hypothesis testing	30
Recommendations and design of A/B test(s) to validate their impact	32
Case Study 2: ETH Wallets	35
Question 1:	35
Question 2:	38

Case Study 1: Expanding a Crypto Bank

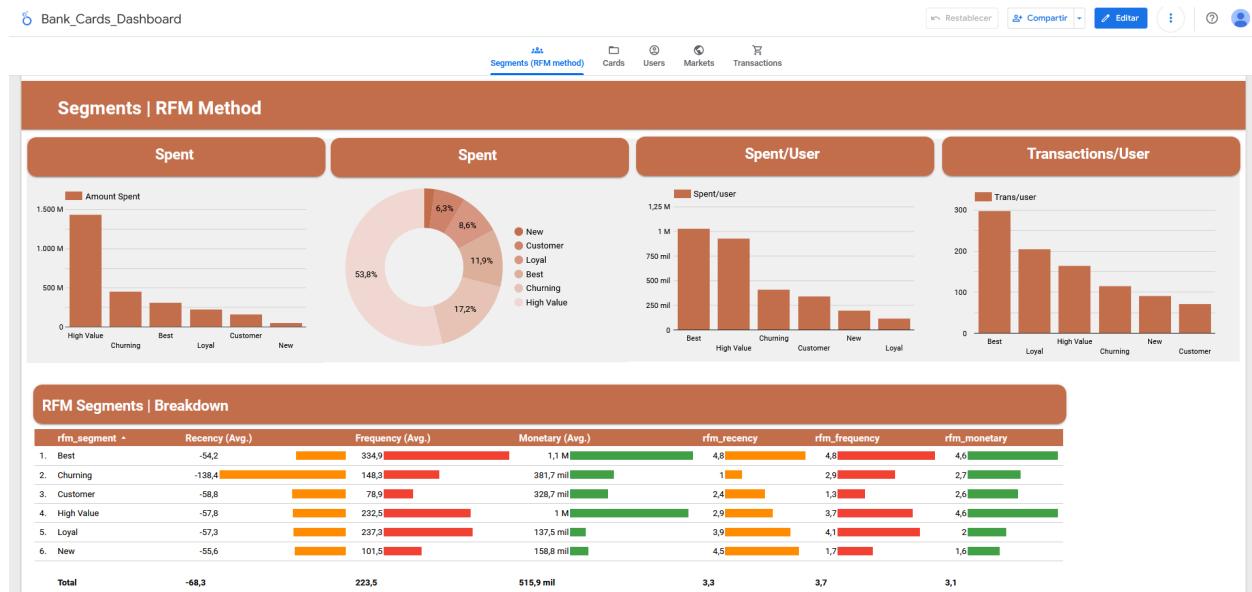
Offering a Global Card

Approach, choices, thought process

1. First of all I started by **visualizing** the data from the different tables to get a picture of the data (**Data Studio** dashboard). Summarizing and categorizing data and analyzing trends and patterns over time.
2. Then I used **BigQuery** to explore the data, by joining the different tables and also including external info (geographical data, economics, banking, crypto, Merchants codes, internet penetration, etc.). Basically I tried to connect the dots and create tables to further analyze in my Dashboard. In this last step I created multiple custom dimensions and categorizations, while also creating calculated metrics.
3. I ended up with lots of **tables and charts** from which I derived lots of interesting insights about the bank's card performance, transaction behavior, user persona, segments, markets and user retention.

Dashboard

⇒ [Link to Data Studio dashboard.](#)



I built a Dashboard with **multiple pages** for each area of analysis:

- Transactions
- Users
- Cards
- Segments
- etc.

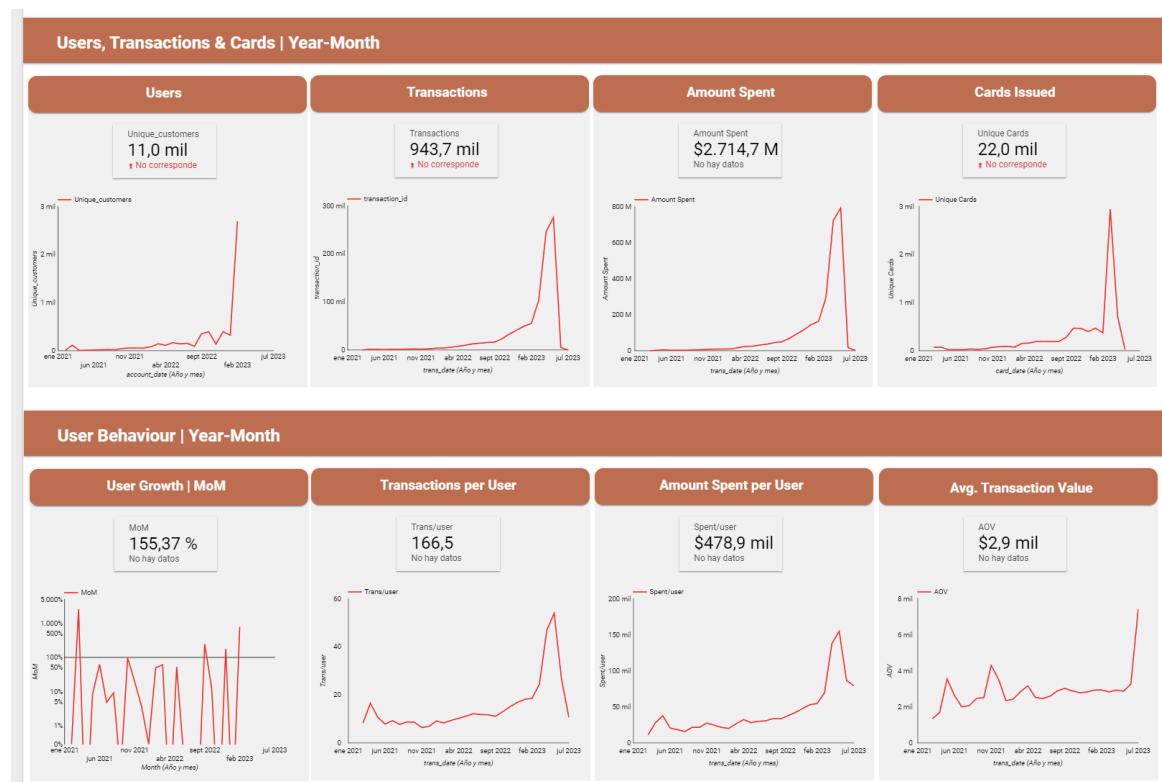
Exploratory data analysis, Interpretation of results and insights

⇒ [Link to folder with all the queries.](#)

General Insights

First I got some general insights that describe the overall performance of the card:

- Customers: 11k
- Transactions: 943K
- Spent: \$2.714M
- Cards issued: 22k
- Trans/user: 166,5
- Spent/user: \$478,9K
- AOV: \$2,9K
- MoM user growth avg: 155%



⇒ [Link to file with queries for Calculated metrics](#)

⇒ [Link to file with queries for MoM](#)

Interpretation of results:

The bank has **grown** exponentially this year (users and spend). However we can clearly see in the charts above that the results are very **weird**. There were huge **spikes** in users, cards and transactions followed by nothing for several months.

- First, **4,9K users** (45% of total user base) created an account on **feb 28, 2023** and after that day nobody else created an account. Of these users, **75%** already had a card previous to account creation (card issuing date before account creation date).

My assumptions from this behavior are that there might be an error in the dataset or maybe people could have a card before an account and one day the bank created an account for all of them.



Users

Date_Dimension_account	Unique_customers	(%)
1. 28 feb 2023	4,9 mil	44,27 %
2. 30 sept 2022	45	0,41 %
3. 30 nov 2022	41	0,37 %
4. 31 dic 2022	39	0,35 %
5. 31 ago 2022	35	0,32 %
6. 11 dic 2022	34	0,31 %
7. 31 ene 2023	33	0,3 %
8. 23 dic 2022	33	0,3 %
Total	11 mil	100 %

1 - 100 / 671 < >

```

9   , date_min as (
10  select customer_id, account_date, min(card_date) as min_card,
11  from users_feb
12  group by customer_id, account_date
13  )
14  )
15
16  , days_before_card as (
17  select customer_id, account_date, min_card, date_diff(min_card, account_date, day) as days_before
18  from date_min
19  )
20  )
21
22
23  select count(distinct customer_id) as users_had_card_before, round(count(distinct customer_id)/(select count(distinct customer_id) from days_before_card),2) as pct
24  from days_before_card
25  where days_before<=0

```

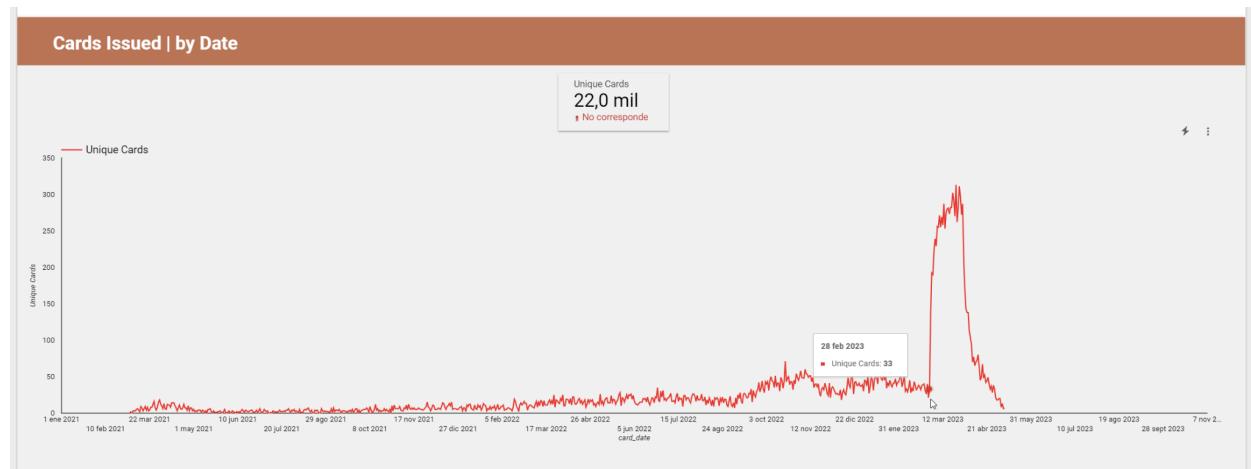
Presiona Alt+F1 para ver las opciones de acc

Resultados de la consulta

INFORMACIÓN DEL TRABAJO **RESULTADOS** GRÁFICO **VISTA PREVIA** JSON DETALLES DE LA EJECUCIÓN GRÁFICO DE EJECUCIÓN

Fila	users_had_card_before	pct
1	3166	0.75

- Second, there is a spike in **cards in March** (36% of all cards), driven by new users acquired in February.



Cards

	Date Dimension_Cards	Unique_cards	(%)
1.	1 mar 2023	8,1 mil	36,79 %
2.	1 abr 2023	2 mil	9,29 %
3.	1 ene 2023	1,4 mil	6,14 %
4.	1 oct 2022	1,3 mil	6,05 %
5.	1 nov 2022	1,2 mil	5,63 %
6.	1 dic 2022	1,1 mil	4,88 %
7.	1 feb 2023	1 mil	4,62 %
8.	1 sept 2022	720	3,28 %
	Total	22 mil	100 %

1 - 27 / 27 < >

- Third, we have a spike in **transactions** and amount spent during **March-May 2023 (67% of all transactions)**. And also as well followed by very few transactions until July, where eventually there were no more transactions.



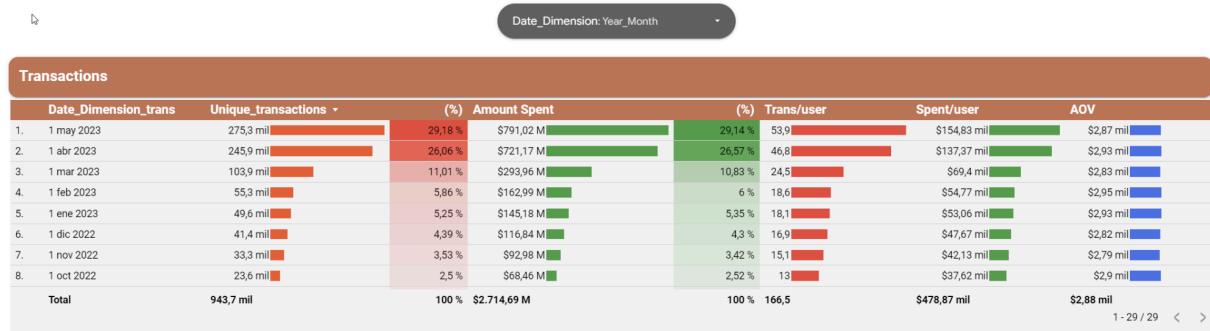
Transactions

Second, I went on to further analyze the **transactions** in greater detail to see if there was anything in particular during the spike period that would explain the behavior. While also trying to analyze changes in user **demography** or **card** type/currency during the spike vs. the average.

Interpretation of results:

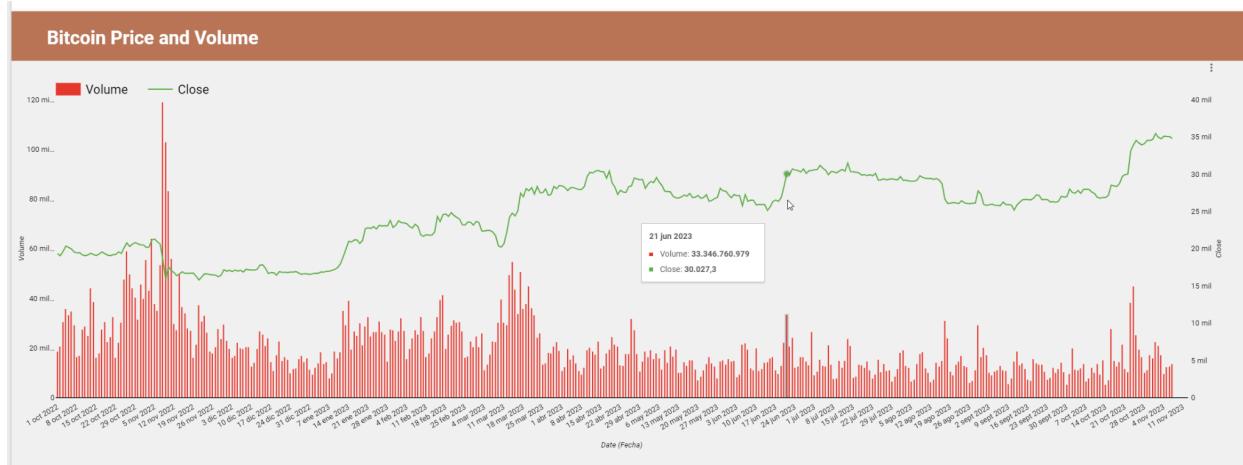
- There were no changes in the transaction variables (payment method, status, type, etc.)
- There was no difference in user demography or card type/currency, etc.
- The only **changes** were:
 - 2x user base increase in February
 - 2X increase in transactions/User and amount spent/user during peak.
 - Spike in **Bitcoin** price (**mar-May**) that might have incentivized people to spend their BTC.

⇒ [Link to file with queries for joining all tables](#)



Date_Dimension_trans	Unique_transactions	(%)	Amount Spent	(%)	Trans/user	Spent/user	AOV
1. 1 may 2023	275,3 mil	29,18 %	\$791,02 M	29,14 %	53,9	\$154,83 mil	\$2,87 mil
2. 1 abr 2023	245,9 mil	26,06 %	\$721,17 M	26,57 %	46,8	\$137,37 mil	\$2,93 mil
3. 1 mar 2023	103,9 mil	11,01 %	\$293,96 M	10,83 %	24,5	\$69,4 mil	\$2,83 mil
4. 1 feb 2023	55,3 mil	5,86 %	\$162,99 M	6 %	18,6	\$54,77 mil	\$2,95 mil
5. 1 ene 2023	49,6 mil	5,25 %	\$145,18 M	5,35 %	18,1	\$53,06 mil	\$2,93 mil
6. 1 dic 2022	41,4 mil	4,39 %	\$116,84 M	4,3 %	16,9	\$47,67 mil	\$2,82 mil
7. 1 nov 2022	33,3 mil	3,53 %	\$92,98 M	3,42 %	15,1	\$42,13 mil	\$2,79 mil
8. 1 oct 2022	23,6 mil	2,5 %	\$68,46 M	2,52 %	13	\$37,62 mil	\$2,9 mil
Total	943,7 mil	100 %	\$2.714,69 M	100 %	166,5	\$478,87 mil	\$2,88 mil

1 - 29 / 29 < >

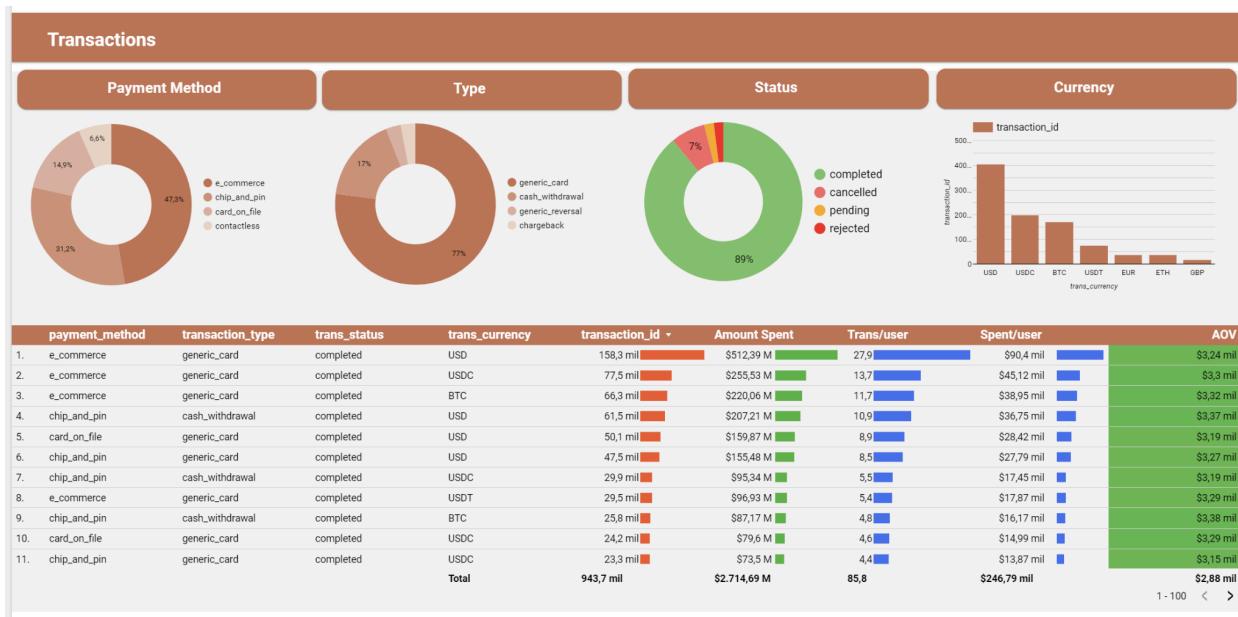


All Tables (joining up the data)

Third, I analyzed the different variables of **user**, **card** and **transaction** data:

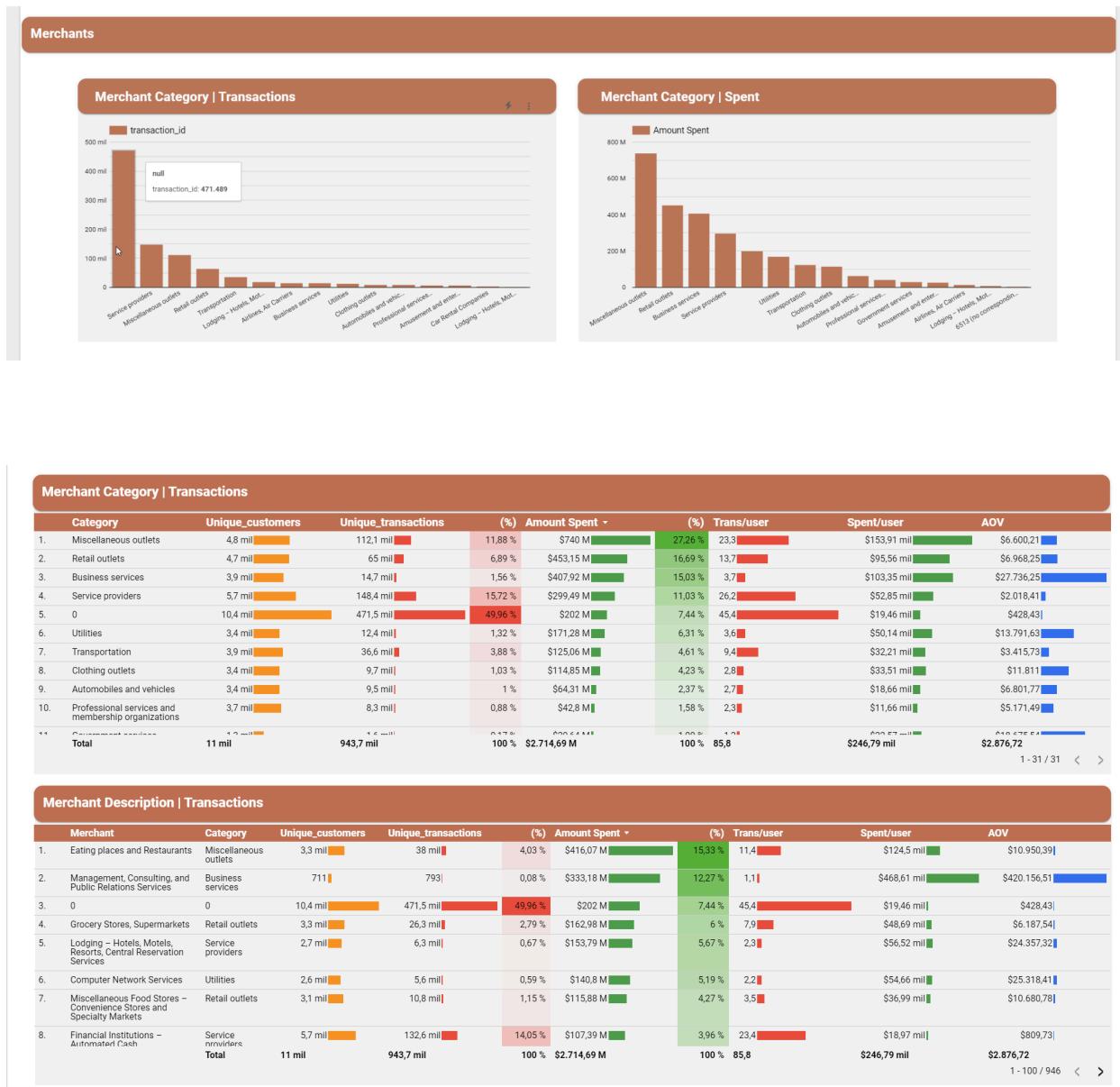
a. Transactions

- Most of the transactions were in ecommerce (47%), using USD (43%)



Transactions Mcc's

- Transactions with **MCCs with wrong values** (50% didn't match neither mcc_codes dataset nor external MCC datasets I downloaded)
- From the transactions with correct MCC values, the **top transactions** by total spending were:
 - ATMs (14%)
 - Restaurants, groceries and bars (12% - Miscellaneous outlets)



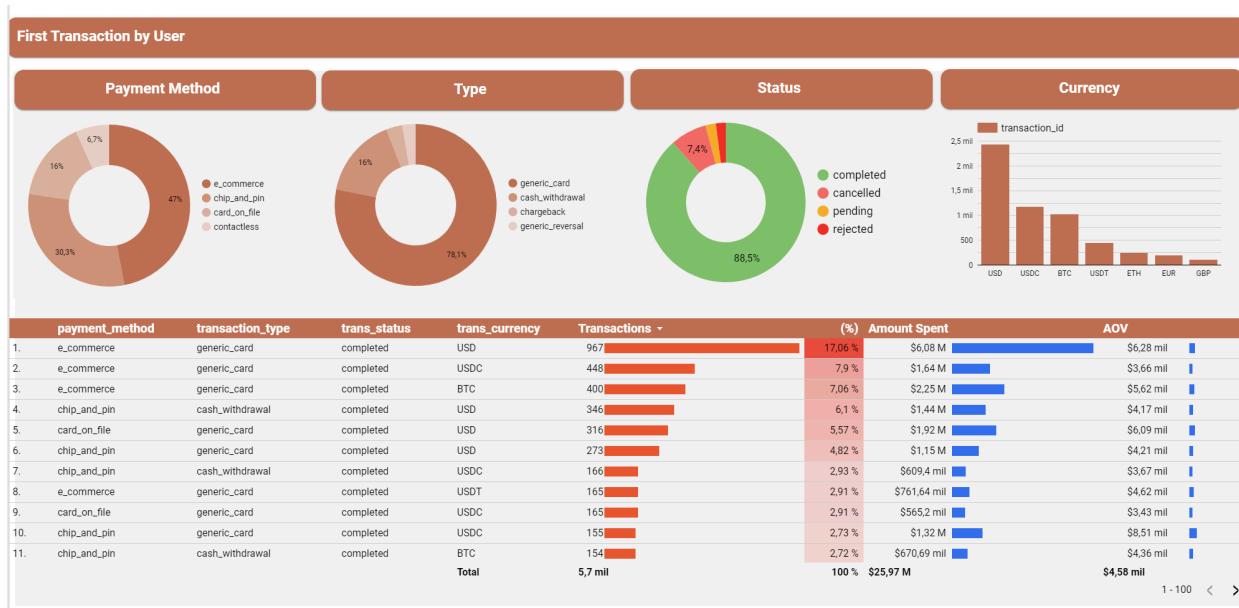
Top 1% Transactions

The biggest transactions by spending (**top1%**) accumulated **\$700M (25%)** of total spent, were \$900k avg and were done solely on **Business services** (Management, Consulting, and Public Relations Services).



First transactions

- The most common **first transactions** made by users were:
 - 17% ecommerce, using USD
 - 17% Restaurants, groceries and bars (Miscellaneous outlets)
 - 15% ATM



⇒ [Link to file with queries for first transactions](#)

Cohort Analysis

Report measuring monthly user retention.

Results: 84% avg. yearly user retention.

Retention Cohorts														
cohort_month (...)	0	1	2	3	4	5	6	7	8	9	10	11	12	cohort_index / user_count
feb 2022	82	81	77	76	73	74	74	73	74	75	75	75	75	-
mar 2022	109	108	101	88	87	87	88	88	90	91	90	92	93	-
abr 2022	161	155	144	132	129	130	130	132	133	132	135	136	140	-
may 2022	150	150	139	119	109	111	113	114	114	116	121	131	131	-
jun 2022	182	178	160	148	144	146	149	150	148	154	158	158	-	-
jul 2022	168	167	153	141	138	140	141	142	148	149	147	3	-	-
ago 2022	180	176	164	152	146	148	150	156	157	156	2	1	-	-
sept 2022	184	180	170	155	154	154	158	162	161	1	-	-	-	-
oct 2022	349	346	324	296	290	299	304	300	5	1	-	-	-	-
nov 2022	407	401	378	347	341	353	353	9	-	-	-	-	-	-
dic 2022	285	279	266	247	249	247	10	2	-	-	-	-	-	-
ene 2023	339	334	318	298	297	4	2	-	-	-	-	-	-	-
feb 2023	279	273	262	247	5	1	-	-	-	-	-	-	-	-
mar 2023	1,246	1,248	1,176	58	2	-	-	-	-	-	-	-	-	-
abr 2023	1,003	942	67	26	-	-	-	-	-	-	-	-	-	-
may 2023	16	4	1	-	-	-	-	-	-	-	-	-	-	-



R24	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	cohort_month	0	1	2	3	4	5	6	7	8	9	10	11	12
2	2021-03	100%	100.0%	88.9%	77.8%	72.2%	66.7%	72.2%	72.2%	72.2%	72.2%	72.2%	72.2%	72.2%
3	2021-04	100%	94.5%	86.8%	84.6%	81.3%	81.3%	78.0%	81.3%	82.4%	83.5%	80.2%	81.3%	-
4	2021-05	100%	93.8%	91.7%	91.7%	93.8%	89.6%	89.6%	93.8%	87.5%	85.4%	87.5%	89.6%	-
5	2021-06	100%	95.2%	95.2%	95.2%	81.0%	81.0%	76.2%	81.0%	76.2%	81.0%	76.2%	85.7%	85.7%
6	2021-07	100%	92.0%	88.0%	84.0%	80.0%	80.0%	84.0%	76.0%	80.0%	80.0%	80.0%	84.0%	84.0%
7	2021-08	100%	97.2%	94.4%	86.1%	75.0%	72.2%	72.2%	77.8%	77.8%	77.8%	77.8%	83.3%	83.3%
8	2021-09	100%	96.2%	80.8%	80.8%	73.1%	73.1%	73.1%	73.1%	73.1%	73.1%	73.1%	73.1%	73.1%
9	2021-10	100%	97.6%	92.9%	90.5%	92.9%	92.9%	92.9%	90.5%	90.5%	92.9%	92.9%	92.9%	92.9%
10	2021-11	100%	96.2%	96.2%	84.6%	82.7%	82.7%	82.7%	84.6%	84.6%	84.6%	84.6%	86.5%	86.5%
11	2021-12	100%	100.0%	94.6%	85.1%	82.4%	83.8%	85.1%	83.8%	82.4%	86.5%	87.8%	86.5%	86.5%
12	2022-01	100%	93.9%	92.7%	85.4%	79.3%	80.5%	81.7%	80.5%	81.7%	85.4%	84.1%	85.4%	-
13	2022-02	100%	98.8%	93.9%	92.7%	89.0%	90.2%	90.2%	89.0%	90.2%	91.5%	91.5%	91.5%	91.5%
14	2022-03	100%	99.1%	92.7%	80.7%	79.8%	80.7%	80.7%	82.6%	83.5%	82.6%	84.4%	85.3%	-
15	2022-04	100%	96.3%	89.4%	82.0%	80.1%	80.7%	80.7%	82.0%	82.6%	82.0%	83.9%	84.5%	87.0%
16	2022-05	100%	100.0%	92.7%	79.3%	72.7%	74.0%	75.3%	76.0%	77.3%	80.7%	87.3%	87.3%	-
17	2022-06	100%	97.8%	87.9%	81.3%	79.1%	80.2%	81.9%	82.4%	81.3%	84.6%	86.8%	86.8%	0.0%
18	2022-07	100%	99.4%	91.1%	83.9%	82.1%	83.3%	83.9%	84.5%	88.1%	88.7%	87.5%	1.8%	0.0%
19	2022-08	100%	97.8%	91.1%	84.4%	81.1%	82.2%	83.3%	86.7%	87.2%	86.7%	1.1%	0.6%	0.0%
20	2022-09	100%	97.8%	92.4%	84.2%	83.7%	83.7%	85.9%	88.0%	87.5%	0.5%	0.0%	0.0%	0.0%
21	2022-10	100%	99.1%	92.8%	84.8%	83.1%	85.7%	87.1%	86.0%	1.4%	0.3%	0.0%	0.0%	0.0%
22	2022-11	100%	98.5%	92.9%	85.3%	83.8%	86.7%	86.7%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%
23	2022-12	100%	97.9%	93.3%	86.7%	87.4%	86.7%	3.5%	0.7%	0.0%	0.0%	0.0%	0.0%	0.0%
24	2023-01	100%	98.5%	93.8%	87.9%	87.6%	1.2%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
25	2023-02	100%	97.8%	93.9%	88.5%	1.8%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
26	2023-03	100%	98.8%	93.3%	4.6%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
27	2023-04	100%	93.9%	8.7%	2.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
28	2023-05	100%	25.0%	6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
29	AVG.	100%	94.6%	85.5%	76.1%	81.9%	81.7%	78.7%	81.9%	82.4%	82.6%	83.0%	84.5%	84.6%
30														

⇒ [Link to file with queries for Cohorts](#)

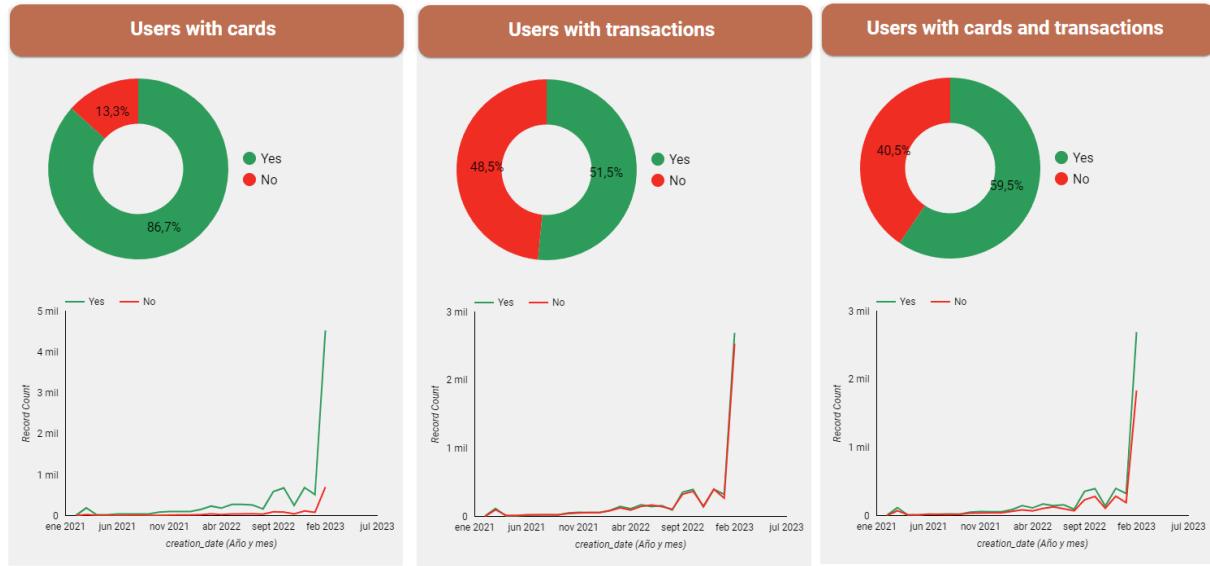
b. Users

- Most of the users are women (61%)
- Aged between 26-36 years (71%)
- With incomes:
 - 50% between \$50-100
 - 40% \$200-250k
- Most common occupations are Crypto trader and blockchain developers (15%)



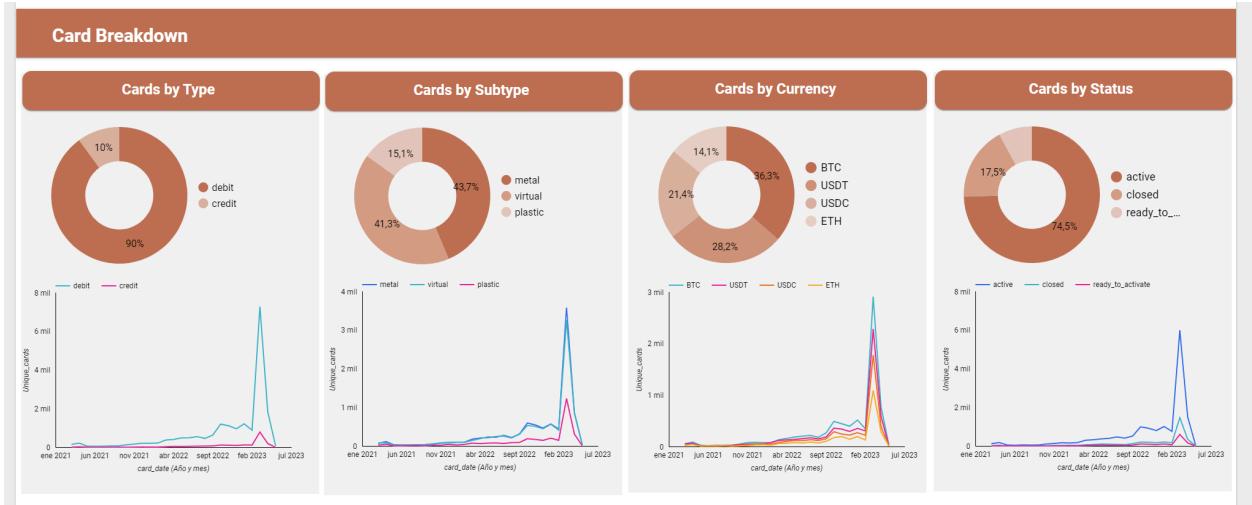
C. Cards

- 87% of users have a card (of whom 60% made transactions)
 - **53%** of total users made transactions
 - 74% of issued cards are active



⇒ [Link to file with queries for cards](#)

- **90%** of cards are **debit** cards
 - Most common subtypes are Metal cards (43%), followed closely by virtual (41%).
 - Most common currency is BTC (36%), despite USD being by far the most common transaction currency.
 - The most common card is debit, BTC, metal (15%)
 - The most common first card by user is BTC, debit, metal (6,32%)



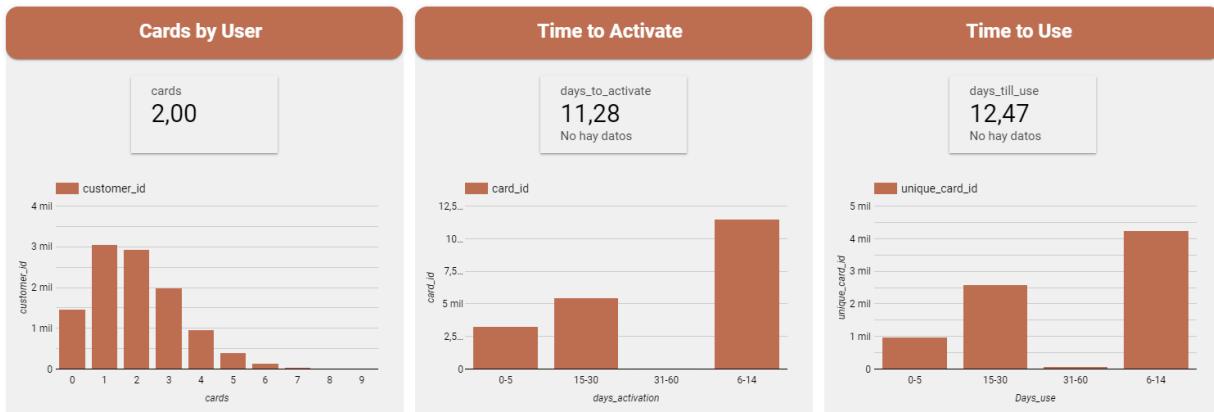
Card Type | User Behaviour

card_currency	card_type	subtype	Cards	(%)	Transactions	(%)	Amount Spent	(%)	AOV	Spent/user	Trans/user
1. BTC	debit	metal	3,2 mil	14,52 %	139,6 mil	14,8 %	\$409,48 M	15,08 %	\$2,93 mil	\$149,17 mil	50,9
2. BTC	debit	virtual	2,9 mil	13,36 %	125,5 mil	13,3 %	\$335,8 M	12,37 %	\$2,68 mil	\$129,55 mil	48,4
3. USDT	debit	metal	2,4 mil	11,03 %	102,6 mil	10,88 %	\$280,79 M	10,34 %	\$2,74 mil	\$129,04 mil	47,2
4. USDT	debit	virtual	2,3 mil	10,5 %	98,9 mil	10,48 %	\$295,7 M	10,89 %	\$2,99 mil	\$142,64 mil	47,7
5. USDC	debit	virtual	1,8 mil	8,13 %	78,5 mil	8,32 %	\$247,64 M	9,12 %	\$3,16 mil	\$150,54 mil	47,7
6. USDC	debit	metal	1,8 mil	8,2 %	73,7 mil	7,81 %	\$213,02 M	7,85 %	\$2,89 mil	\$127,71 mil	44,2
7. ETH	debit	metal	1,2 mil	5,54 %	56,3 mil	5,96 %	\$159,11 M	5,86 %	\$2,83 mil	\$137,76 mil	48,7
8. ETH	debit	virtual	1,1 mil	5,23 %	49,5 mil	5,24 %	\$140,4 M	5,17 %	\$2,84 mil	\$129,05 mil	45,5
9. BTC	debit	plastic	1,1 mil	4,82 %	45,7 mil	4,85 %	\$149,67 M	5,51 %	\$3,27 mil	\$148,04 mil	45,2
Total			22 mil	100 %	943,7 mil	100 %	\$2,714,69 M	100 %	\$2,88 mil	\$284,74 mil	99

Card Type | What Type users get first?

card_rank	currency	type	subtype	Cards	(%)
1. 1	BTC	debit	metal	1,4 mil	6,32 %
2. 1	BTC	debit	virtual	1,3 mil	5,74 %
3. 1	USDT	debit	metal	1 mil	4,72 %
4. 1	USDT	debit	virtual	984	4,48 %
5. 2	BTC	debit	metal	939	4,27 %
Total				22 mil	100 %

- On avg. **each user:**
 - o has 2 cards
 - o Takes 11 days to activate it
 - o Takes 12 days to make a transaction with it



Segments

I analyzed possible user segments in **2 ways**:

- 1) “**RFM**” Method: recency, frequency and monetary value (quartiles)
- 2) **Spending** Segments (“Whales” Method)

By using these types of segmentations we can understand the different **kinds** of users we have (good, bad, etc.). Analyze their behavior, characteristics and plan our actions towards them accordingly.

1) RFM

By “**RFM**” Segments, analyzing **recency, frequency and monetary value (quartiles)**

- Best: top quartile on each metric
- High Value: Customers in the top quartile for monetary value and frequency and any quartile for Recency.
- Loyal: Customers in the top quartile for frequency and any quartile for monetary value and Recency.
- Churning: Customers in the bottom quartile for Recency and any quartile for frequency and monetary value.
- New: Customers in the top quartile for Recency and any quartile for frequency and monetary value.

By analyzing by **RFM**, we can **know**:

- who are our **new** clients
- who are **churning** and try to reactivate them (emails, sms, push notifications, discounts, coupons, etc.)
- The behavior of our **loyal and high value** clients and how to retain them even longer, and increase their frequency and value.
- Who are our **best** customers and give them a special treatment while also understanding them to get more like them.

Results:



Best customer Segment:

- Demographic indicators similar to avg
- 95% more credit cards than avg.
- Spending categories
 - 15% restaurants
 - 13% en Business services, Management, Consulting, and Public Relations Services

⇒ [Link to file with queries for RFM segments](#)

Hypothesis testing

I tested the “Best” RFM segment. I got all the users from this segment that made purchases and their accumulated value by user and compared it with a random sample (same size) of all the users, to test if there was a difference on the mean spending by user.

Results: Best segment difference in spending was **Statistically significant**.

The screenshot shows a spreadsheet titled "Hypothesis Testing". The data is organized into several columns: A, B, C, D, E, F, G, H, I, J, K. Row 1 contains column headers like "total_spent", "Sample", "total_spent", "size", "mean", "STD", "Sample users", "I=Best", and "Best". Rows 2 through 12 contain numerical values for each of these metrics across 12 different rows. Row 6 is highlighted in blue and contains the formula "T-test (p-value)" in cell I6, which has a yellow background. The value in cell I6 is "0.000004 <0.05", with the "**<0.05**" part highlighted in blue. To the right of this cell, the text "statistically significant" is written in blue. The rest of the table is in grey, and the overall interface looks like a standard spreadsheet application.

	A	B	C	D	E	F	G	H	I	J	K
1	total_spent		Sample		total_spent			Sample users	I=Best	Best	
2	444745.3074		444745.3074		683275.2842		size	306	306		
3	369331.8404		369331.8404		1018349.208		mean	457,666.48	1,051,867.58		
4	75156.14		75156.14		2007167.673		STD	562,249.37	550,671.25		
5	1486830.335		1486830.335		1040429.828						
6	443206.6177		443206.6177		1561184.055		T-test (p-value)	0.000004 <0.05	statistically significant		
7	61386.52		61386.52		591412.8187		alpha	0.05			
8	683734.2798		683734.2798		765468.1678						
9	91188.68		91188.68		764776.9798		2 sample t-test	2 different groups with different variance and unknown population variance			
10	485571.4059		485571.4059		2221734.506						
11	413441.3534		413441.3534		1257489.245						
12	1377875.548		1377875.548		1221655.512						

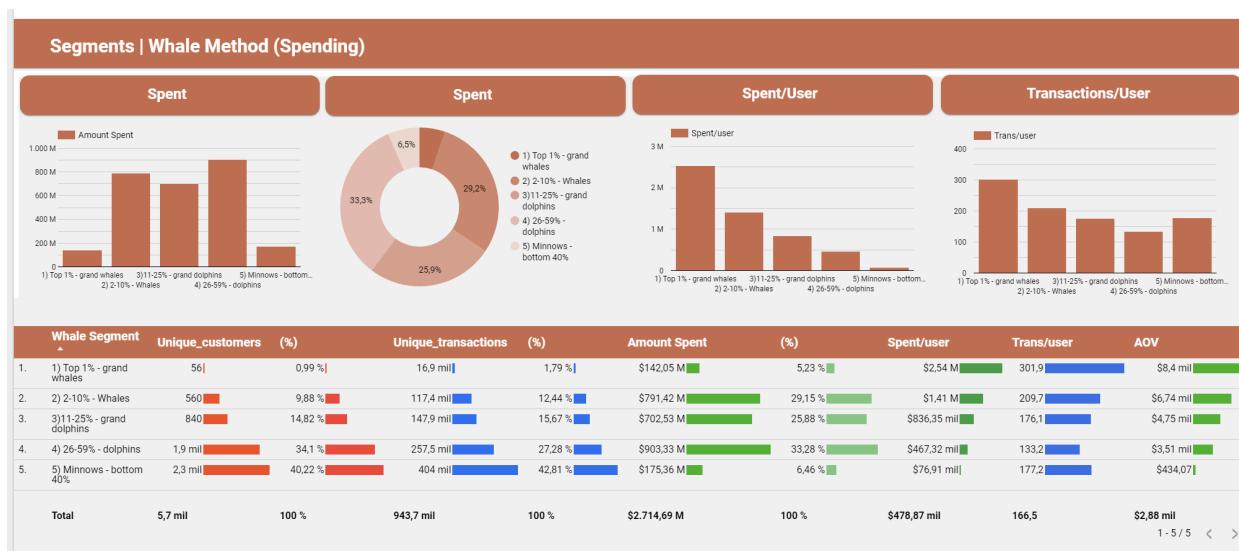
⇒ [Link to file with Hypothesis Testing](#)

2) Spending Segments (“Whales” Method):

- top 1%: grand whales
- 2 - 10%: whales
- 11 - 25%: grand dolphins
- 26 - 60%: dolphins
- less than 60%: minnows.

By analyzing the value users are providing to our business, we can get to know who are the most profitable ones. And by analyzing them, we can understand them better in order to find more customers like them, retain them longer and make them spend even more.

Results:



Top 10% (all whales) of spenders:

- represented 35% of all amount spent
- Spent 3-4 times more than avg.
- 2,5x AOV
- Demographic indicators similar to avg
- 70% more credit cards than avg.
- Spending categories
 - 30% of spending in Business services, Management, Consulting, and Public Relations Services
 - 13% restaurants

⇒ [Link to file with queries for spending segment](#)

Markets

Fifth, I analyzed the Markets and potential for expansion in **2 ways:**

- 1) **Entering** new markets (we don't have users there)
- 2) **Expanding** current markets where we have good users

1) Entering new markets (we don't have users there):

I gathered all relevant countries where we don't have users and analyzed them according to certain **indicators (economic, crypto, banking, internet penetration)**.

- Crypto ownership (%)
- GDP/capita
- Inflation (higher incentive to use USD or crypto than local currency)
- Internet penetration (%)
- Bancarization (people with bank accounts (%))
- Population
- Tax rate (crypto and USD offshore account)



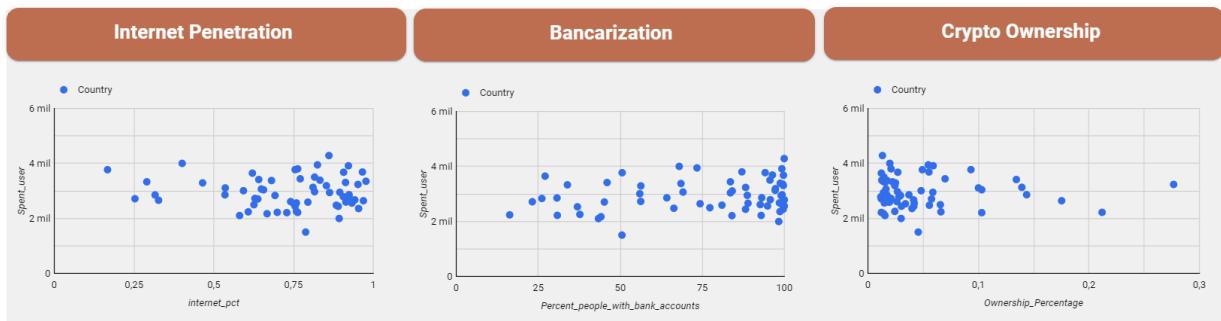
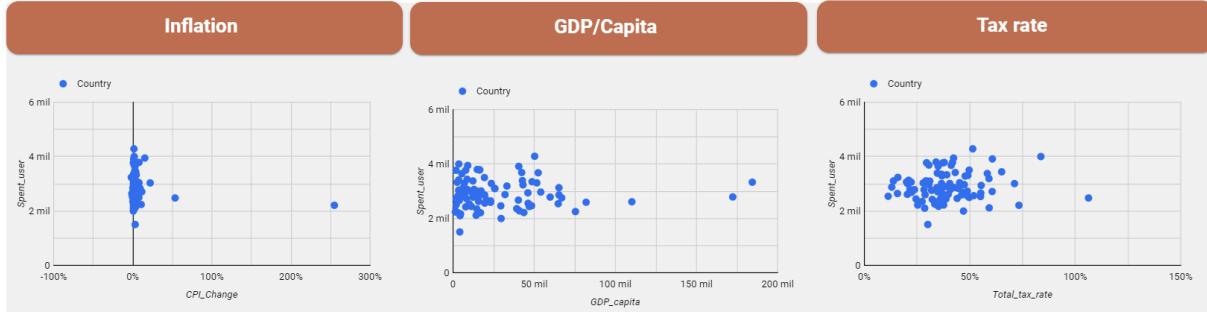
- The best markets according to this analysis were:

- Iran
- Morocco
- Belarus
- Georgia
- Moldova
- Kazakhstan
- Myanmar
- Tunisia
- Ghana

⇒ [Link to file with queries for missing countries](#)

Some of these countries might have **problems** regarding: politics, sanctions, regulations, etc. (i.e: Iran).

Furthermore, I analyzed the **relationship** between card **performance (spent/user)** and the countries **indicators** mentioned above and I found no relationship whatsoever. My assumption is that the sample data is too small and also because it might be dummy data it might not reflect real life behavior.



2) Expanding current markets where we have good users:

The **user** count is distributed around the **world**:

- 50% of users are from Americas, 35% from USA and 15% from LATAM
- 28% from Europe
- 15% Asia
- The remaining are from Oceania and Africa.

The absolut main market right now is the USA (3,8k users, 35%) and the rest of the countries have max between 100-200 users.

Transaction behavior by **Region**

Geo Dimension: region									
Markets breakdown									
Geo Dimension	Unique_customers	(%)	transaction_id	(%)	Amount Spent	(%)	Trans/user	Spent/user	AOV
1. Americas	2,9 mil	50,87 %	475,3 mil	50,36 %	\$1,379,06 M	50,8 %	164,8	\$478,18 mil	\$2,9 mil
2. Europe	1,6 mil	28,74 %	273,4 mil	28,97 %	\$801,67 M	29,5 %	167,8	\$492,13 mil	\$2,93 mil
3. Asia	876	15,45 %	147,3 mil	15,61 %	\$389,45 M	14,3 %	166,1	\$444,58 mil	\$2,64 mil
4. Africa	208	3,58 %	34,9 mil	3,69 %	\$104,55 M	3,9 %	171,7	\$515,05 mil	\$3 mil
5. Oceania	77	1,36 %	12,9 mil	1,36 %	\$39,95 M	1,5 %	166,9	\$518,81 mil	\$3,11 mil

Transaction behavior by **sub-region**

Geo Dimension: Sub_region									
Markets breakdown									
Geo Dimension	Unique_customers	(%)	transaction_id	(%)	Amount Spent	(%)	Trans/user	Spent/user	AOV
1. Northern America	2 mil	35,37 %	329,4 mil	34,9 %	\$939,99 M	34,6 %	164,3	\$468,82 mil	\$2,85 mil
2. Latin America and the Caribbean	879	15,51 %	145,9 mil	15,46 %	\$439,08 M	16,2 %	166	\$499,52 mil	\$3,01 mil
3. Northern Europe	529	9,33 %	89,9 mil	9,53 %	\$271,27 M	10,0 %	170	\$512,79 mil	\$3,02 mil
4. Southern Europe	502	8,86 %	83,9 mil	8,89 %	\$223,95 M	8,2 %	167,2	\$446,12 mil	\$2,67 mil
5. Western Asia	381	6,72 %	63,2 mil	6,7 %	\$170,22 M	6,3 %	165,9	\$446,77 mil	\$2,69 mil
6. Western Europe	304	5,36 %	51,3 mil	5,44 %	\$158,49 M	5,8 %	168,8	\$521,34 mil	\$3,09 mil
7. Eastern Europe	294	5,10 %	48,3 mil	5,12 %	\$147,96 M	5,5 %	164,3	\$503,28 mil	\$3,06 mil
8. South-eastern Asia	191	3,37 %	33,5 mil	3,55 %	\$91,25 M	3,4 %	175,2	\$477,73 mil	\$2,73 mil
9. Sub-Saharan Africa	171	3,02 %	28,4 mil	3 %	\$86,13 M	3,2 %	165,8	\$503,69 mil	\$3,04 mil
10. Eastern Asia	170	3 %	29,4 mil	3,11 %	\$76,81 M	2,8 %	172,7	\$451,83 mil	\$2,62 mil
11. Southern Asia	134	2,36 %	21,3 mil	2,25 %	\$51,17 M	1,9 %	158,6	\$381,9 mil	\$2,41 mil

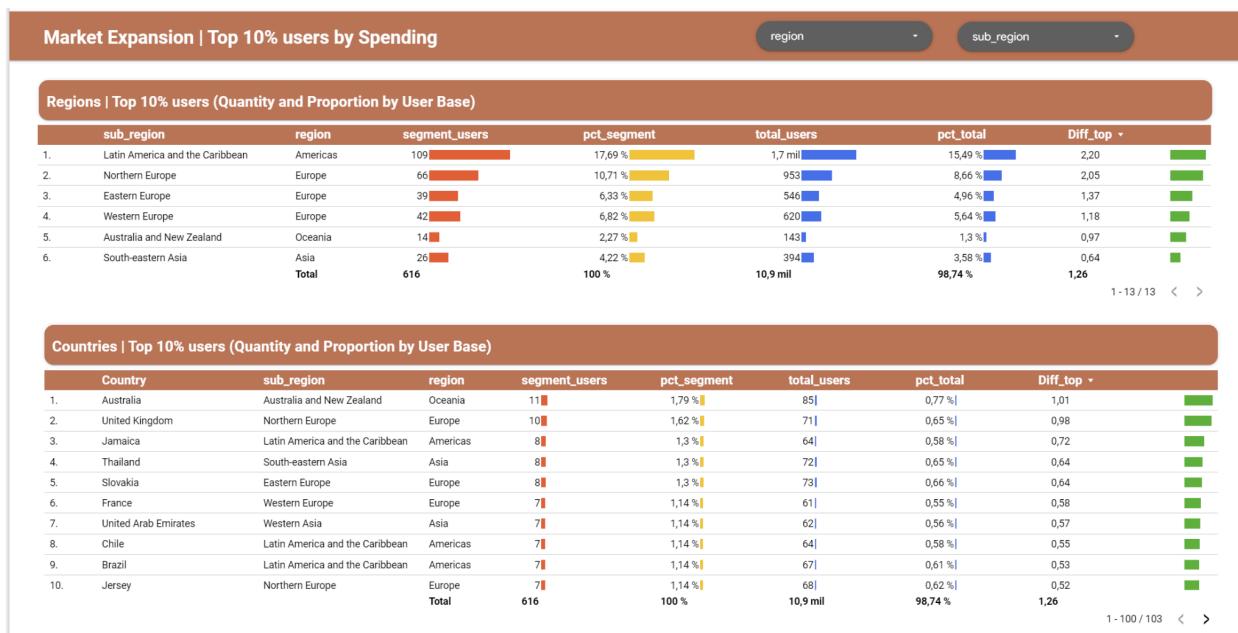
Transaction behavior by **Country**

Geo Dimension: Country									
Markets breakdown									
Geo Dimension	Unique_customers	(%)	transaction_id	(%)	Amount Spent	(%)	Trans/user	Spent/user	AOV
1. United States of America	2 mil	34,87 %	324,6 mil	34,4 %	\$928,47 M	34,2 %	164,2	\$469,64 mil	\$2,86 mil
2. United Kingdom of Great Britain and Northern Ireland	43	0,76 %	8,6 mil	0,91 %	\$31,67 M	1,2 %	199,8	\$736,57 mil	\$3,69 mil
3. Martinique	57	1,01 %	9,4 mil	0,99 %	\$28,5 M	1,0 %	164,1	\$500 mil	\$3,05 mil
4. Australia	47	0,83 %	8,6 mil	0,91 %	\$25,4 M	0,9 %	182,1	\$540,44 mil	\$2,97 mil
5. Serbia	46	0,81 %	8,3 mil	0,88 %	\$24,38 M	0,9 %	181,5	\$530,03 mil	\$2,92 mil
6. Isle of Man	47	0,83 %	6,8 mil	0,72 %	\$24,28 M	0,9 %	144,9	\$516,5 mil	\$3,57 mil
7. Slovakia	39	0,69 %	6,9 mil	0,73 %	\$24,1 M	0,9 %	176,5	\$618,04 mil	\$3,5 mil
8. Jersey	38	0,67 %	7,3 mil	0,77 %	\$23,83 M	0,9 %	192,2	\$627,14 mil	\$3,26 mil
9. Brazil	41	0,72 %	6,8 mil	0,72 %	\$23,37 M	0,9 %	165,8	\$570 mil	\$3,44 mil
10. Netherlands	37	0,65 %	6,1 mil	0,64 %	\$22,29 M	0,8 %	163,8	\$602,37 mil	\$3,68 mil
11. Thailand	39	0,69 %	5,9 mil	0,62 %	\$22,19 M	0,8 %	150,8	\$568,99 mil	\$3,77 mil
12. Trinidad and Tobago	45	0,79 %	7,4 mil	0,78 %	\$22,12 M	0,8 %	163,4	\$491,59 mil	\$3,01 mil

High spending segment (Countries with most of this users)

By Spending Segments (Whale Method):

- **Top 10%** spending users (35% of total spending) and the proportion each **country** has of them compared to their participation in the user base.
- The **best markets** according to this analysis were:
 - Australia
 - UK
 - Jamaica
 - Thailand
 - United Arab Emirates

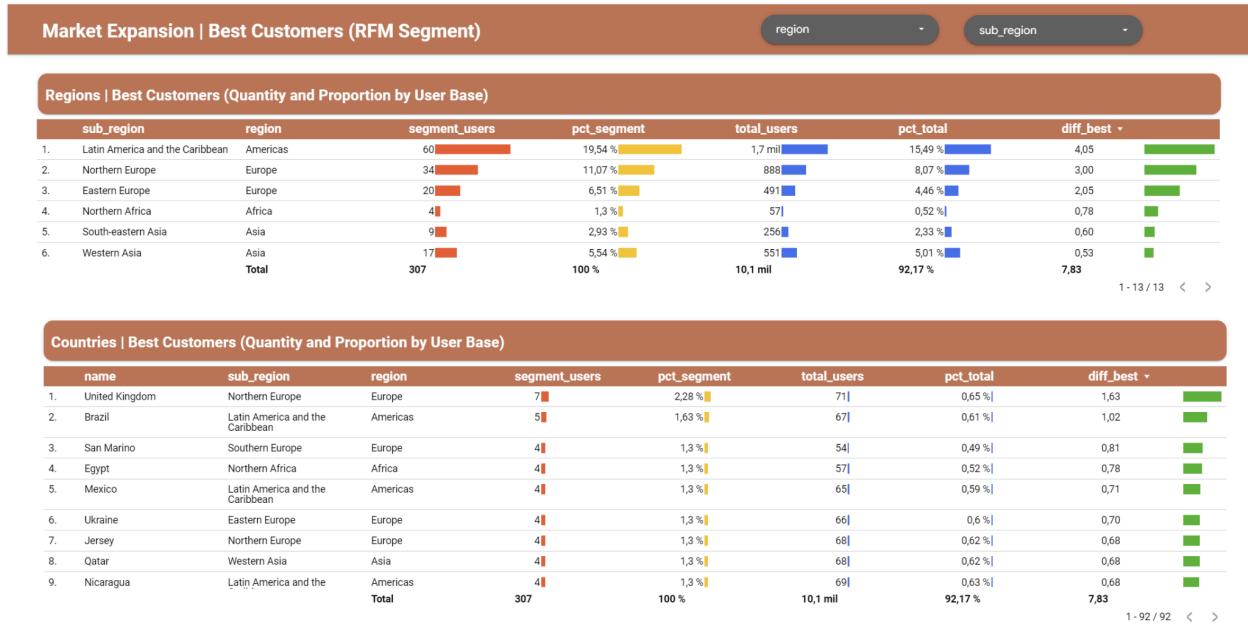


⇒ [Link to file with queries for top 10% spenders and country data](#)

RFM

By **Best customer Segment (RFM Method)**, those with high recency, frequency and monetary value (top quartile on each metric)

- Analyzing the proportion each **country** has of them compared to their participation in the user base.
- The **best markets** according to this analysis were:
 - UK
 - Brazil
 - Egypt
 - Mexico
 - San Marino
 - Ukraine

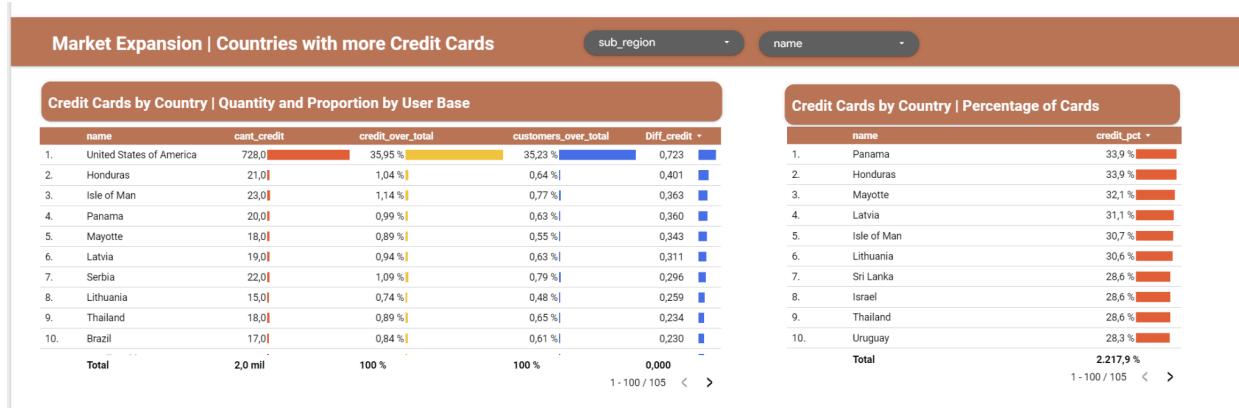


⇒ [Link to file with queries to analyze the proportions by market](#)

Credit cards

Countries with the most **credit cards** (volume and percentage of users).

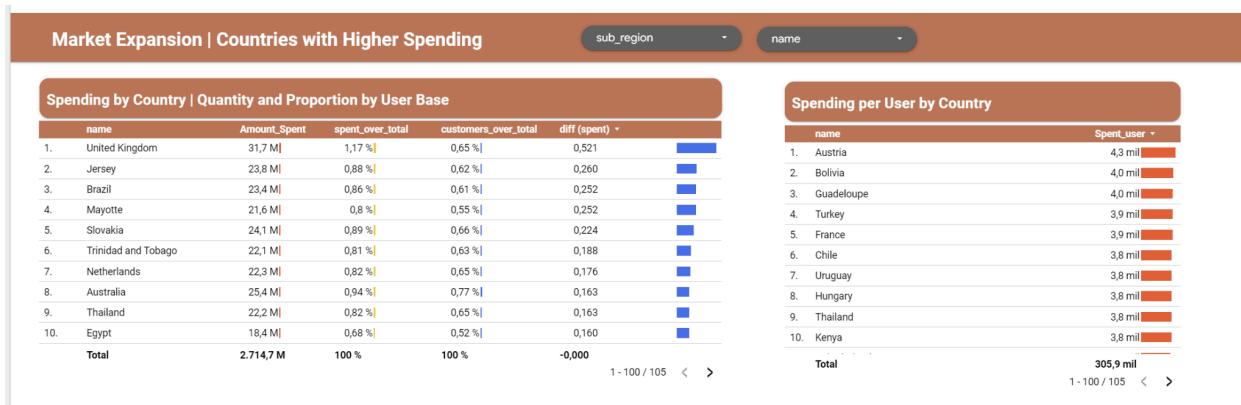
- Because credit cards normally are **more profitable** than debit cards (interests and higher fees)
- The best markets according to this analysis were:
 - USA
 - Honduras
 - Panama
 - Mayotte
 - Latvia
 - Lithuania



⇒ [Link to file with queries for top Countries with the most credit cards](#)

Biggest spending

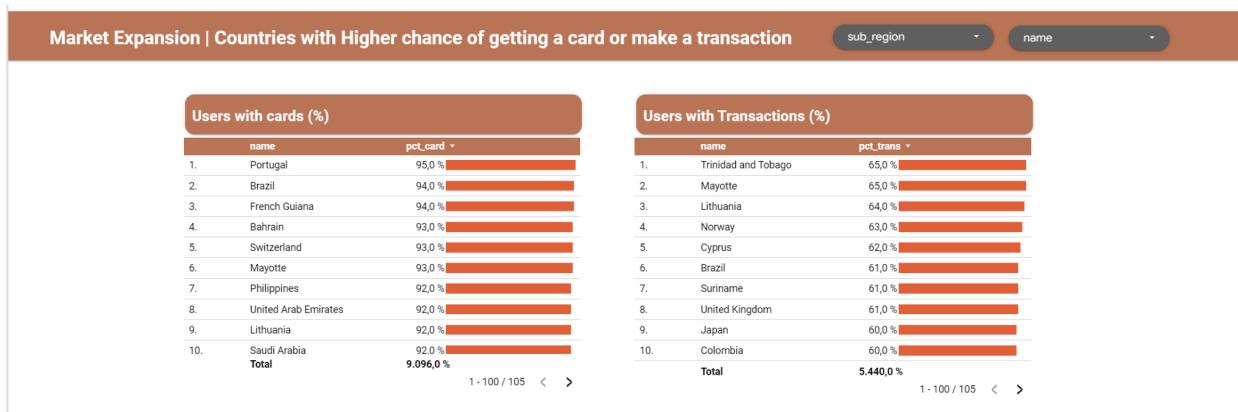
- Countries with **biggest spending** in proportion to users
 - The best markets according to this analysis were:
 - UK
 - Brazil
 - Mayotte
 - Slovakia



Highest ratios of users with cards and transactions (%):

Countries with **highest ratios** of users with **cards and transactions (%)**:

- The best markets according to this analysis were:
 - UK
 - Brazil
 - Mayotte
 - Lithuania



⇒ [Link to file with queries for Countries with highest ratios of users with cards and transactions \(%\):](#)

Hypothesis testing

I tested the **United Kingdom** and **Brazil**, which were countries that appeared a lot in my market analysis as potential good markets with good users.

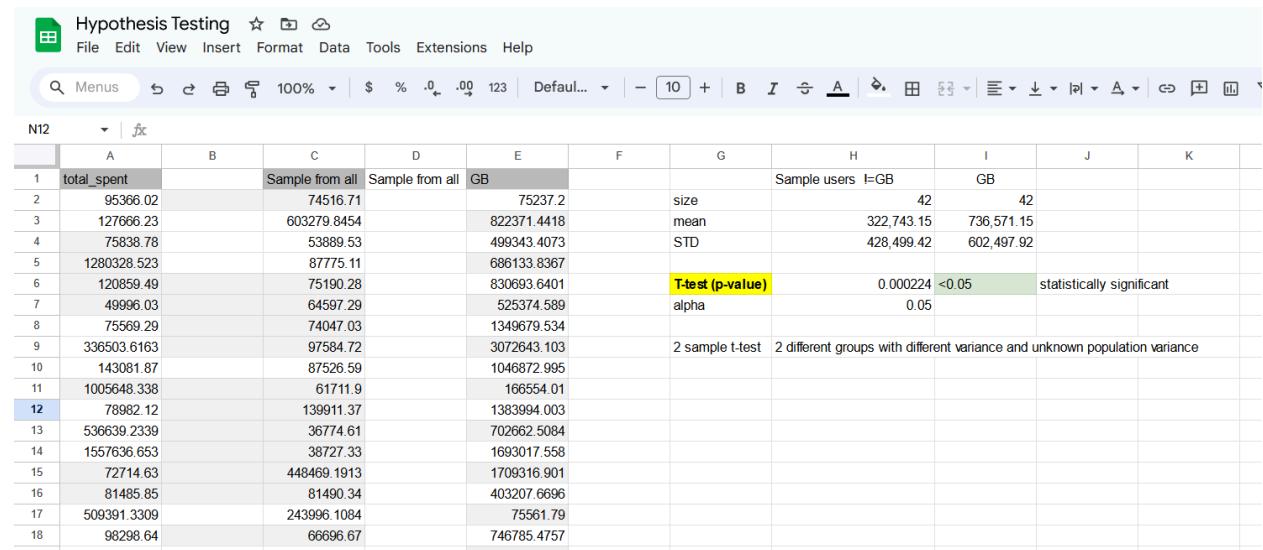
Also I tested **Spain**, one of the **worst** in terms of spending. Just in case to test if my method I applied to the Hypothesis testing was correct.

In my method I got all the users from UK/GB that made purchases and compared it with a random sample (same size) of all the users worldwide (except UK) to test if there was a difference on the mean spending by user.

I used a **t-test**, with 1 tail (increase) and 2 samples with different variances.

Results: UK and Brazil differences in spending were Statistically significant and Spain not.

UK:



The screenshot shows a spreadsheet titled "Hypothesis Testing" with various tabs like File, Edit, View, Insert, Format, Data, Tools, Extensions, and Help. The main table has columns labeled A through K. Row 1 contains column headers: "total_spent", "Sample from all", "Sample from all", "GB", "size", "mean", "STD", "Sample users", "I=GB", "GB", "J", and "K". Rows 2 through 18 contain data for individual users. Row 12 highlights the results of a two-sample t-test for GB vs. I=GB, showing a p-value of 0.000224, which is less than 0.05, indicating statistical significance. The notes column (J) states: "2 sample t-test 2 different groups with different variance and unknown population variance".

N12	A	B	C	D	E	F	G	H	I	J	K
1	total_spent	Sample from all	Sample from all	GB	size	mean	STD	Sample users	I=GB	GB	
2	95366.02		74516.71		75237.2				42	42	
3	127666.23		603279.8454		822371.4418				322,743.15	736,571.15	
4	75838.78		53889.53		499343.4073				428,499.42	602,497.92	
5	1280328.523		87775.11		686133.8367						
6	120859.49		75190.28		830693.6401			T-test (p-value)	0.000224	<0.05	statistically significant
7	49996.03		64597.29		525374.589			alpha	0.05		
8	75569.29		74047.03		1349679.534						
9	336503.6163		97584.72		3072643.103			2 sample t-test	2 different groups with different variance and unknown population variance		
10	143081.87		87526.59		1046872.995						
11	1005648.338		61711.9		166554.01						
12	78982.12		139911.37		1383994.003						
13	536639.2339		36774.61		702662.5084						
14	1557636.653		38727.33		1693017.558						
15	72714.63		448469.1913		1709316.901						
16	81485.85		81490.34		403207.6696						
17	509391.3309		243996.1084		75561.79						
18	98298.64		66696.67		746785.4757						

BR:

	A	B	C	D	E	F	G	H	I
1	Random Sample		BR			Sample users I=BR	BR		
2	67985.96		415568.7767		size	40	40		
3	68637.9		1200412.299		mean	303,674.78	570,000.29		
4	737603.2451		1837554.877		STD	296,549.91	487,693.43		
5	60647.9		1203329.817						
6	64120.04		579932.2902		T-test (p-value)	0.001972 <0.05	statistically significant		
7	484118.1578		585857.4764		alpha	0.05			
8	71818.6		545012.6388						
9	73986.82		128915.11		2 sample t-test	2 different groups with different variance and unknown population variance			
10	68975.99		335801.0872						
11	60647.9		579932.2902						

ES:

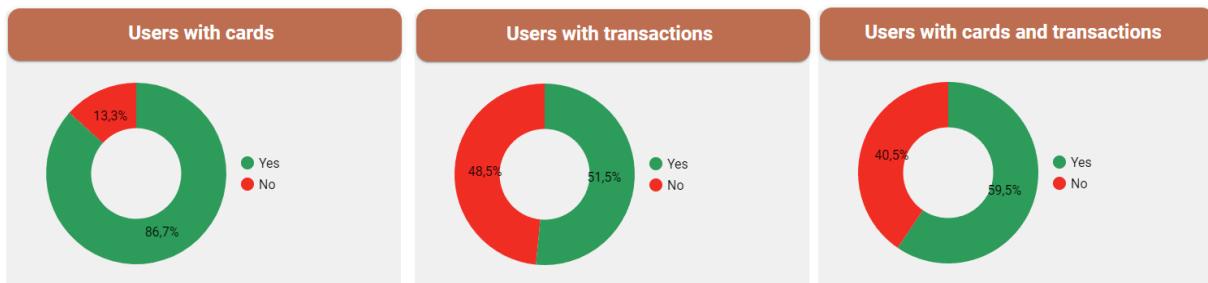
Hypothesis Testing									
	A	B	C	D	E	F	G	H	I
1	total_spent		total_spent			Sample users I=ES	ES		
2	95366.02		500597.8092		size	36	36		
3	127666.23		85646.19		mean	361,319.69	345,064.16		
4	75838.78		75442.08		STD	411,826.95	286,387.07		
5	1280328.523		692065.4732						
6	120859.49		94333.26		T-test (p-value)	0.422178 >0.05	not statistically significant		
7	49996.03		434015.054		alpha	0.05			
8	75569.29		62958.01						
9	336503.6163		506955.8296		2 sample t-test	2 different groups with different variance and unknown population variance			
10	143081.87		763512.9057						
11	1005648.338		134554.59						
12	78982.12		75838.78						
13	536639.2339		576611.5641						
14	1557636.653		581585.5366						
15	72714.63		596768.85						
16	81485.85		91689.31						
17	509391.3309		501226.1639						
18	98298.64		99792.03						

⇒ [Link to file for Hypothesis testing](#)

Recommendations and design of A/B test(s) to validate their impact

1) Increase the amount of users that make transactions.

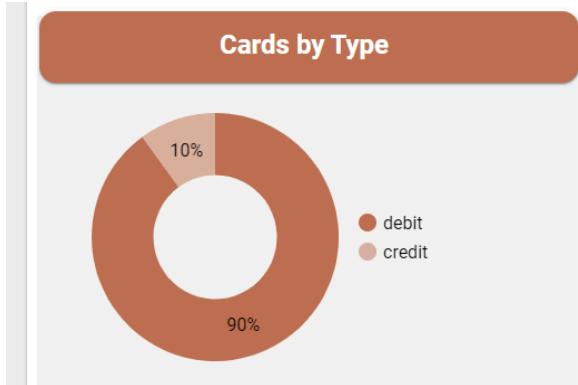
- As we saw on my analysis:
 - Only 53% of total users made transactions
 - Only 60% of users with cards made transactions



- We should be **targeting** those that **didn't transact** with discounts, coupons or any incentive for the most common MCCs for first transactions (17% Restaurants, groceries and bars (Miscellaneous outlets))
- **A/B test:**
 - Significance level (**alpha**): 0.05
 - Grab all the users that didn't transact and created an account recently (year=2023)
 - Take **2 samples** (size: we don't know the population variance nor delta, so we cannot set an exact size. but I would take the 2.8k users from this year that didn't transact and take 2 random samples of **1.000** users)
 - **Control** sample: we wont send them anything
 - **Test** sample: send emails, sms, push notifications, ads, etc
 - Run the experiment for **1 month** and analyze the results by measuring either:
 - The difference in mean amount of transactions between the 2 samples. (T-test)
 - The difference in mean transaction rate (%) (transact/all). (Bernoulli Z-test)
 - If the results are **statistically significant (p-value < 0.05 (alpha))**, we can reject the null hypothesis (that both samples performed the same) and assume that our campaign worked and users that were **targeted transacted more** than those who weren't.

2) Increase the amount of credit cards:

- Credit cards normally are **more profitable** than debit cards (interests and higher fees)
- As we saw on my analysis:
 - **90%** of cards are **debit** cards



- We should be **targeting** those that **don't have credit cards** to incentivise them to order one. We could send emails, sms, push notifications and advertising that make our credit cards attractive (i.e: **perks and benefits** like one-time signing bonus, cash backs, rewards points, flier miles, etc.)
- **A/B test:**
 - Significance level (**alpha**): 0.05
 - Grab all the users that don't have a credit card and created an account recently (year=2023)
 - Take **2 samples** (size: we don't know the population variance nor delta, so we cannot set an exact size. but I would take the 7k users from this year that don't have a credit card and take 2 samples of **2-3k** users.
 - **Control** sample: we wont send them anything
 - **Test** sample: send emails, sms, push notifications, ads, etc
 - Run the experiment for **1 month** and analyze the results by measuring either:
 - The difference in mean amount of credit cards ordered between the 2 samples. (T-test)
 - The difference in mean credit cards rate (%) (transact/all). (Bernoulli Z-test)
 - If the results are **statistically significant (p-value < 0.05 (alpha))**, we can reject the null hypothesis (that both samples performed the same) and assume that our campaign worked and users that were **targeted ordered more credit cards** than those who weren't.

3) Reactivate churning users:

- The cost of **acquiring** new customers is **5x higher** than the cost of **retaining** existing

customers. It's crucial to retain our customers over time, and maximize their lifecycle and **LTV**.

- As we saw on my analysis:

- 20% of customers are **churning**, representing 17% of all value.



- We should be **targeting** those users that **didn't transact** in a **while (lower quartile in recency)** with discounts, coupons or any incentive to try to reactivate them.

- A/B test:**

- Significance level (**alpha**): 0.05
- Grab all the users that didn't transact in a while (lower quartile in recency) and created an account recently (year=2023)
- Take **2 samples** (size: we don't know the population variance nor delta, so we cannot set an exact size. but I would take the 600 users from this year that are churning and take 2 samples of **300** users).
 - Control** sample: we wont send them anything
 - Test** sample: send emails, sms, push notifications, ads, etc
- Run the experiment for **1 month** and analyze the results by measuring either:
 - The difference in mean amount of transactions between the 2 samples. (T-test)
 - The difference in mean reactivation rate (%) (transact/all). (Bernoulli)

- Z-test)
- If the results are **statistically significant** (**p-value < 0.05** (alpha)), we can reject the null hypothesis (that both samples performed the same) and assume that our campaign worked and users that were **targeted transacted more** than those who weren't.
- 

Case Study 2: ETH Wallets

Question 1:

Which are the **top 10% addresses** in terms of transaction volume in the past 6 months? What can you tell about their transactional behavior? Classify them in spenders and accumulators.

Results:

[Google Drive Folder with the results in CSV's.](#)

Queries:

```

with

-- table with all the addresses that sent ETH and their total sent value in the last 6
months

sender_addresses AS (
SELECT from_address, sum(value) as Sending_Volume
FROM `bigquery-public-data.crypto Ethereum.transactions`
where date_trunc(date(block_timestamp), MONTH) >= date_trunc(date_add(CURRENT_DATE,
INTERVAL -6 MONTH), MONTH) -- after current month - 6 months (not date - 6 months)
group by (from_address)
)

-- table with all the addresses that received ETH and their total received value in
the last 6 months

,receiver_addresses AS (
SELECT to_address, sum(value) as Receiving_Volume
FROM `bigquery-public-data.crypto Ethereum.transactions`
where date_trunc(date(block_timestamp), MONTH) >= date_trunc(date_add(CURRENT_DATE,
INTERVAL -6 MONTH), MONTH)
group by (to_address)
)

-- table merging 2 previous tables with balances table, showing how much each address
sent and received and calculating new columns: total volume (sent or received) and net
volume (received - sent)

,all_addresses AS (
SELECT address, Sending_Volume,Receiving_Volume,
       Sending_Volume + Receiving_Volume as total_volume,
       Receiving_Volume - Sending_Volume as net_volume
FROM `bigquery-public-data.crypto Ethereum.balances` as a join sender_addresses as s
      ON a.address = s.from_address
      join receiver_addresses as r ON a.address = r.to_address
)

```

```

-- top 10% addresses in terms of transaction volume in the past 6 months. Query
separating table into 10 groups according to their total volume and getting only the
10th decile (biggest group, top 10%)

,top_volume AS (
select *
from (select *, NTILE(10) over(order by total_volume) AS v_decile
      from all_addresses)
where v_decile = 10
order by total_volume DESC
)

-- top 10% of addresses that received more ETH than what they sent (from the top 10%
addresses in transaction volume)

,top_accumulators AS (
select *
from (select *, NTILE(10) over(order by net_volume) AS a_decile
      from top_volume)
where a_decile = 10
order by net_volume DESC
)

-- top 10% of addresses that sent more ETH than what they received (smallest group,
bottom 10%), (from the top 10% addresses in transaction volume)

,top_spenders AS (
select *
from (select *, NTILE(10) over(order by net_volume) AS s_decile
      from top_volume)
where s_decile = 1
order by net_volume ASC
)

/*
-- query to get top 10% addresses in terms of transaction volume in the past 6 months
SELECT *

```

```
from top_volume

-- query to get top 10% spending addresses
SELECT *
from top_spenders

-- query to get top 10% accumulating addresses
SELECT *
from top_accumulators

*/
```

Question 2:

Which are the **top 5% wallets** by balance? For each of them determine:

- first transaction date
- last transaction date
- monthly transacted volume in ETH
- transaction frequency
- total number of transactions
- total transacted volume in ETH

Results:

[Google Drive Folder with the results in CSV's.](#)

I limited the amount of data in the queries and tables because it was too large and took too long to load (i.e: limit 200 wallets and limit 1000 transactions).

Queries:

```
with

-- table with the top 5% wallets by balance (Getting the biggest from 20 groups).

top_5pcnt AS (
  select *
  from (select *, NTILE(20) over(order by eth_balance) AS percentile
        from `bigquery-public-data.crypto Ethereum.balances`)
  where percentile = 20
  order by eth_balance DESC
)

-- table with the transactions made by the top 5% wallets

,transactions AS (SELECT address, from_address, to_address, block_timestamp, value
                  FROM top_5pcnt as a join
bigquery-public-data.crypto Ethereum.transactions as b
                  ON a.address = b.from_address
                  OR a.address = b.to_address
)

/*
-- first and last transaction dates by address (both sending or receiving)

select address, min(block_timestamp) as first_transaction_date, max(block_timestamp)
as last_transaction_date
from transactions
group by address

-- monthly transacted volume in ETH by address and month, (both sending or
receiving), 1 ETH = 10^18 Wei

select address, date_trunc(date(block_timestamp), MONTH) AS month, sum
(value)/10000000000000000 as monthly_transacted_volume_ETH
from transactions
group by address, month
```

```
order by address, month

-- monthly transaction frequency per address

select address, date_trunc(date(block_timestamp), MONTH) AS month, count (address) as freq_transactions_per_month
from transactions
group by address, month
order by address, month

-- total number of transactions by address

select address, count (address) as total_transactions
from transactions
group by address
order by total_transactions desc

-- total transacted volume in ETH by address, 1 ETH = 10^18 Wei

select address, sum (value)/1000000000000000000000000 as total_transacted_volume_ETH
from transactions
group by address
order by total_transacted_volume_ETH desc

*/
```