Mini-Quest: Consultas SQL + Análisis

Adentrándonos en el mundo de las queries en SQL

Proyecto en 2 partes

Parte 1: SQL

Queries en SQL
En donde trabajamos con operaciones lógicas, operaciones numéricas y con strings, agregaciones...

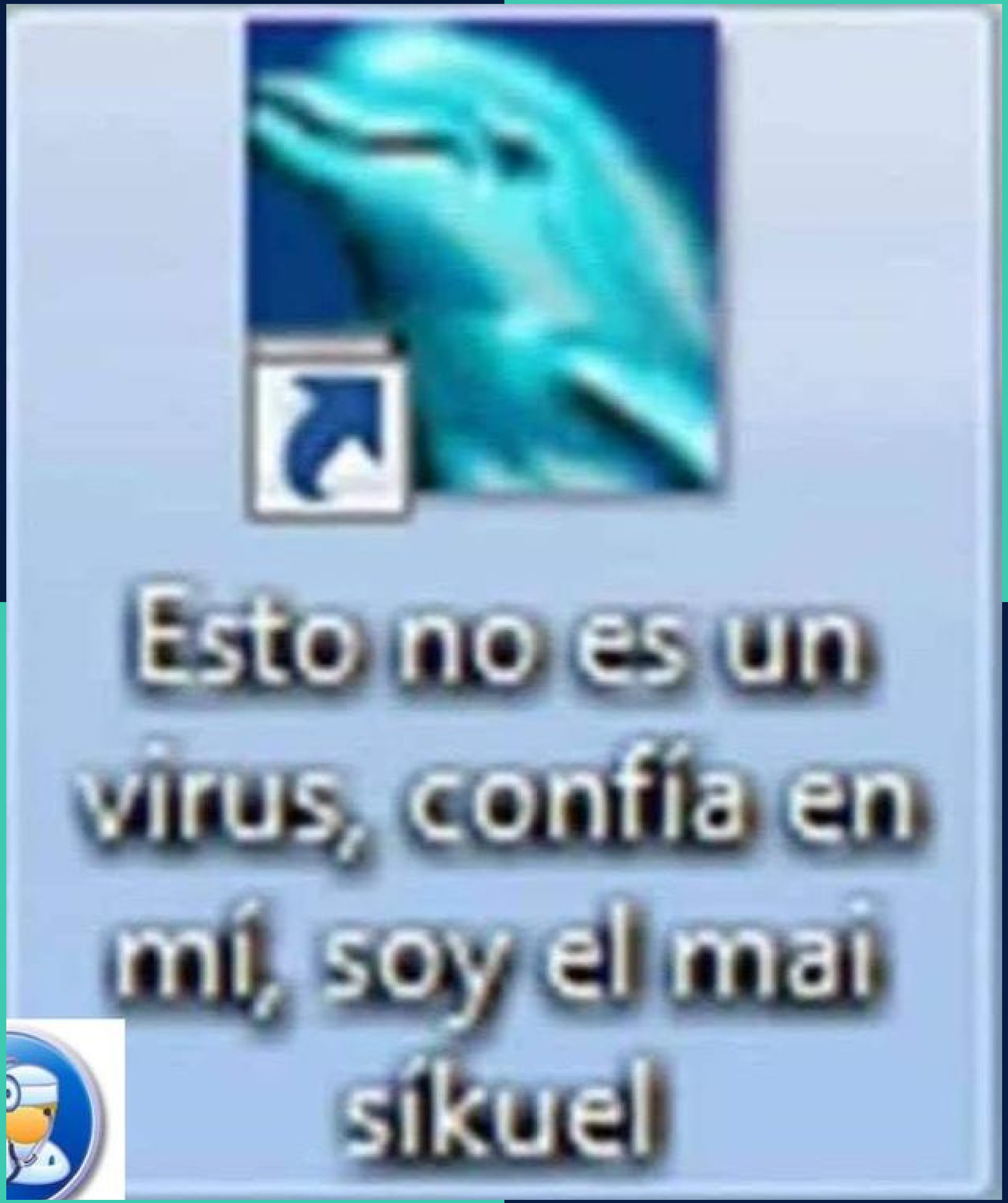
También entramos con trabajos de agrupación: joins, group bys, order bys...

Parte 2: EDA

EDA en Pyhon

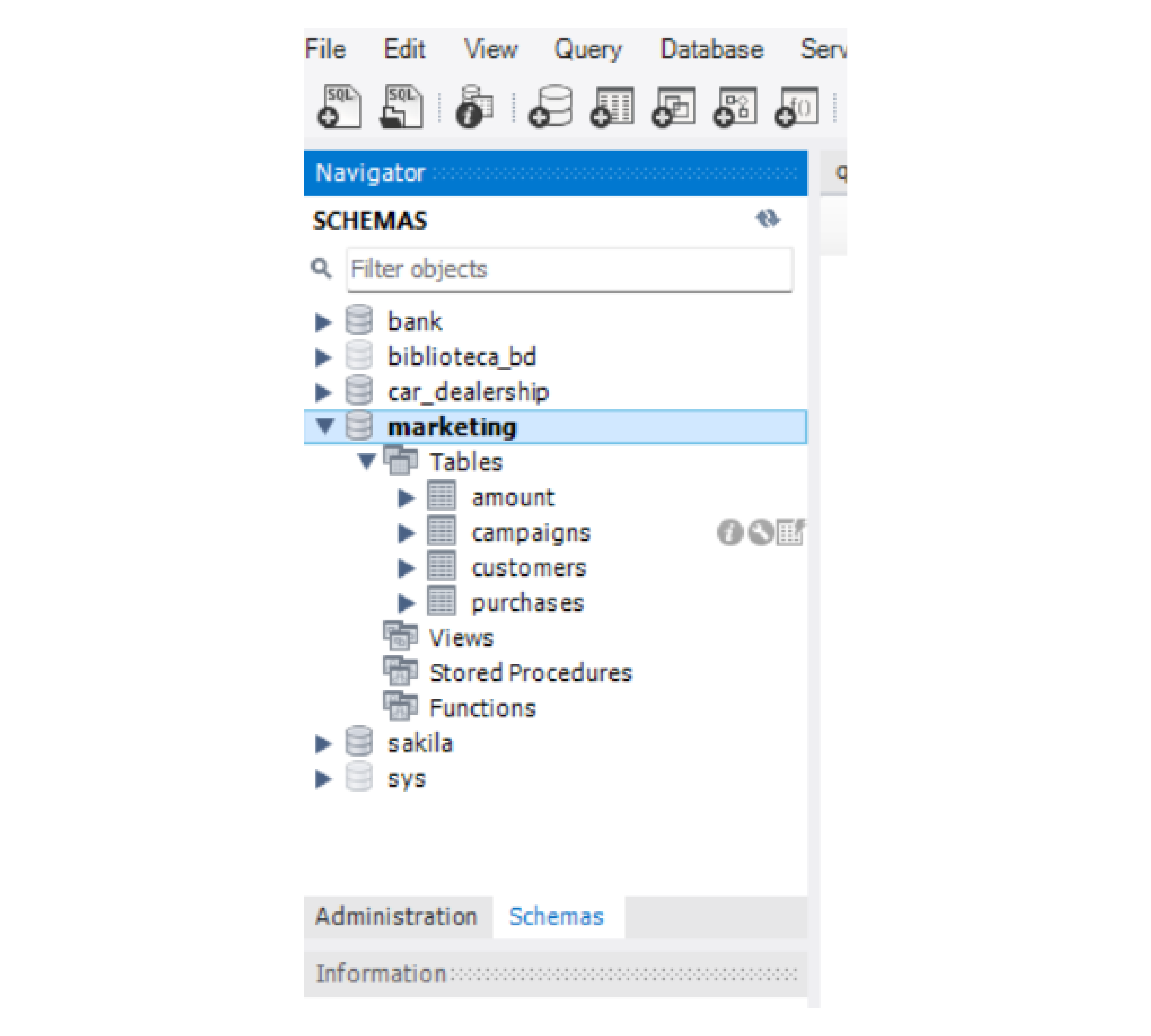
Ploteamos algunas interpretaciones de los CSV importados a través de las queries por SQL

QUERIES



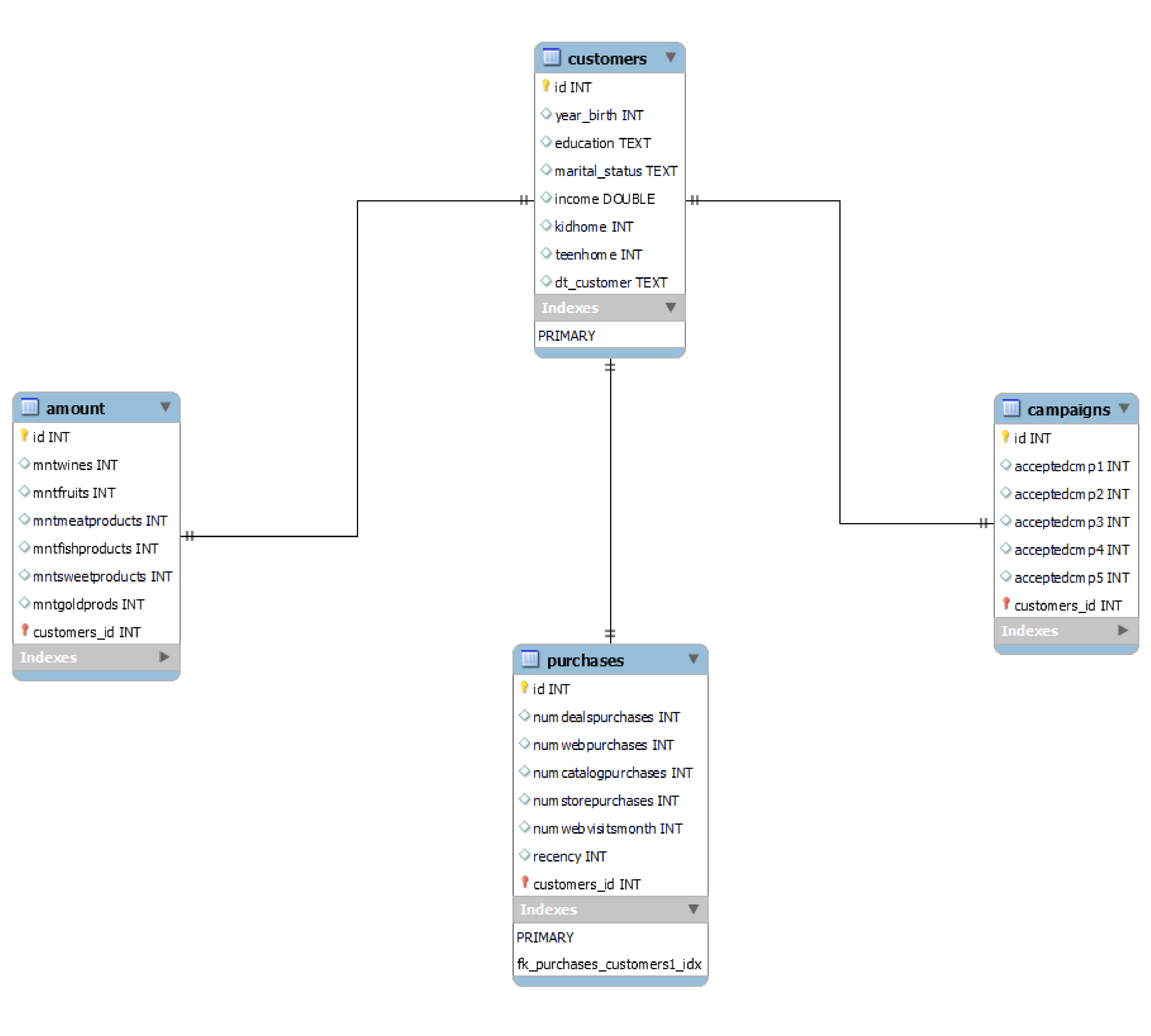
Importamos DB y tablas

A través de Table **Data Import Wizard**



Trazamos diagrama EER

Con **Reverse Engineer** visualizamos todo el esquema con el que trabajaremos



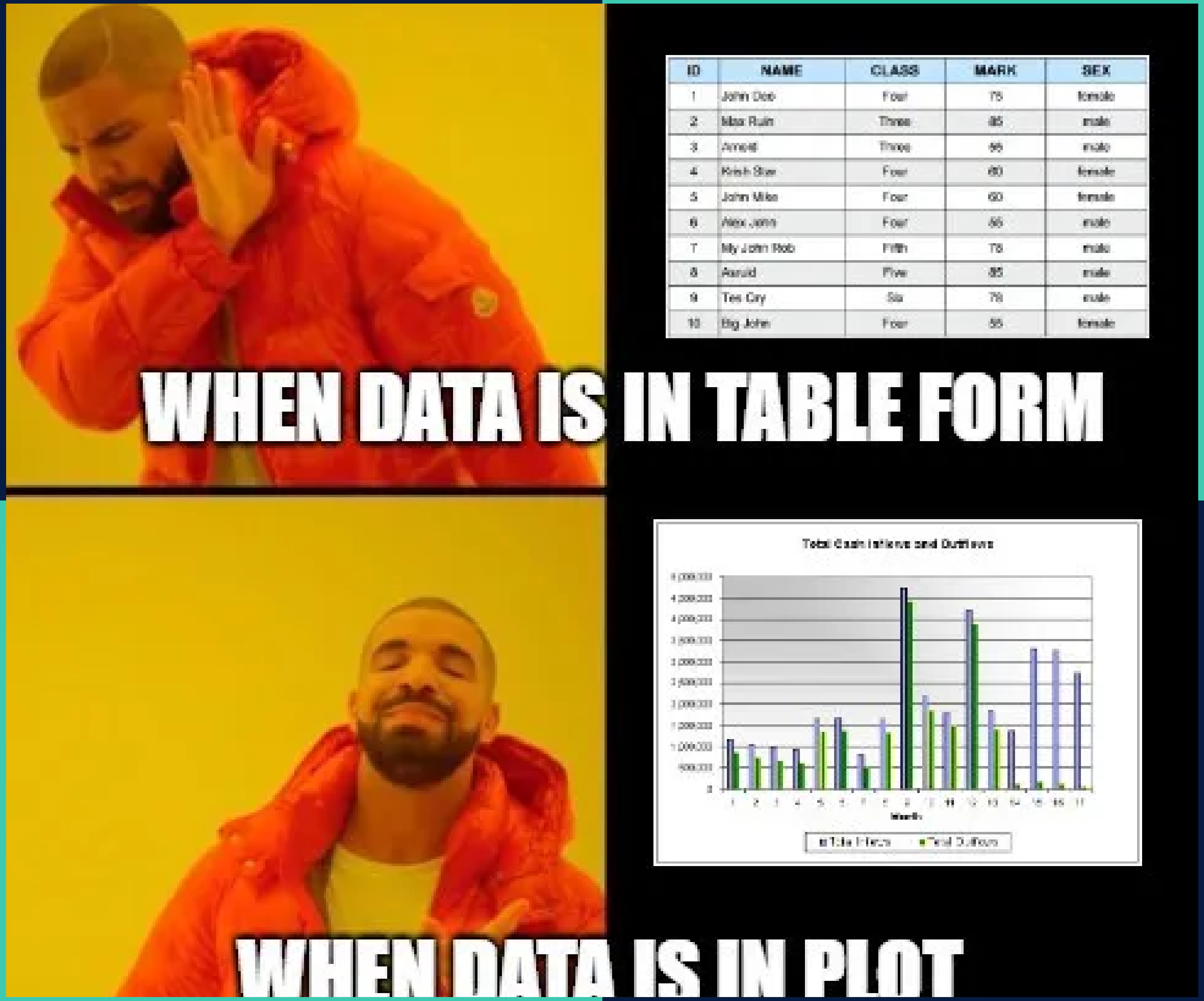
Trabajamos con tablas individualmente

```sql
############################
### CUSTOMERS  #############
############################

-- AVERAGE INCOME PER EDUCATION LEVEL
SELECT education,
       COUNT(id) AS n_of_clients,
       ROUND(AVG(2024 - year_birth),0) AS average_age,
       ROUND(AVG(income),1) AS average_income,
       ROUND(AVG(kidhome),2) AS average_kids_at_home,
       ROUND(AVG(teenhome),2) AS average_teens_at_home
FROM customers
GROUP BY education
ORDER BY n_of_clients DESC;

-- AVERAGE INCOME PER MARITAL STATUS
SELECT marital_status,
    COUNT(id) AS n_of_clients,
    ROUND(AVG(2024 - year_birth),0) AS average_age,
    ROUND(AVG(income),1) AS average_income,
    ROUND(AVG(kidhome),2) AS average_kids_at_home,
    ROUND(AVG(teenhome),2) AS teens_at_home
```

Y con varias tablas unidas a través de Join

```sql
-- AMOUNT SPENT AND PURCHASES PER MARITAL STATUS
SELECT cus.marital_status,
       COUNT(cus.id) AS n_of_clients,
       SUM(amo.mntwines) AS total_wine,
       SUM(amo.mntfruits) AS total_fruit,
       SUM(amo.mntmeatproducts) AS total_meat,
       SUM(amo.mntfishproducts) AS total_fish,
       SUM(amo.mntsweetproducts) AS total_sweet,
       SUM(amo.mntgoldprods) AS total_gold,
       SUM(amo.mntwines + amo.mntfruits + amo.mntmeatproducts + amo.mntfishproducts + amo.mntsweetproducts + amo.mntgoldprods) AS total_all_products,
       SUM(pur.numwebpurchases) AS web_purchases,
       SUM(pur.numcatalogpurchases) AS catalog_purchases,
       SUM(pur.numstorepurchases) AS store_purchases,
       SUM(pur.numwebpurchases + pur.numcatalogpurchases + pur.numstorepurchases) AS total_purchases
FROM customers AS cus
JOIN amount AS amo
ON cus.id = amo.id
JOIN purchases AS pur
ON amo.id = pur.id
WHERE marital_status <> "Absurd" AND marital_status <> "YOLO"
GROUP BY cus.marital_status
ORDER BY n_of_clients DESC;
```

EDA


ID NAME CLASS MARK SEX
WHEN DATA IS IN TABLE FORM
WHEN DATA IS IN PLOT

Ploteamos en Python

De los datasets exportados por las queries en SQL

Data Extraction

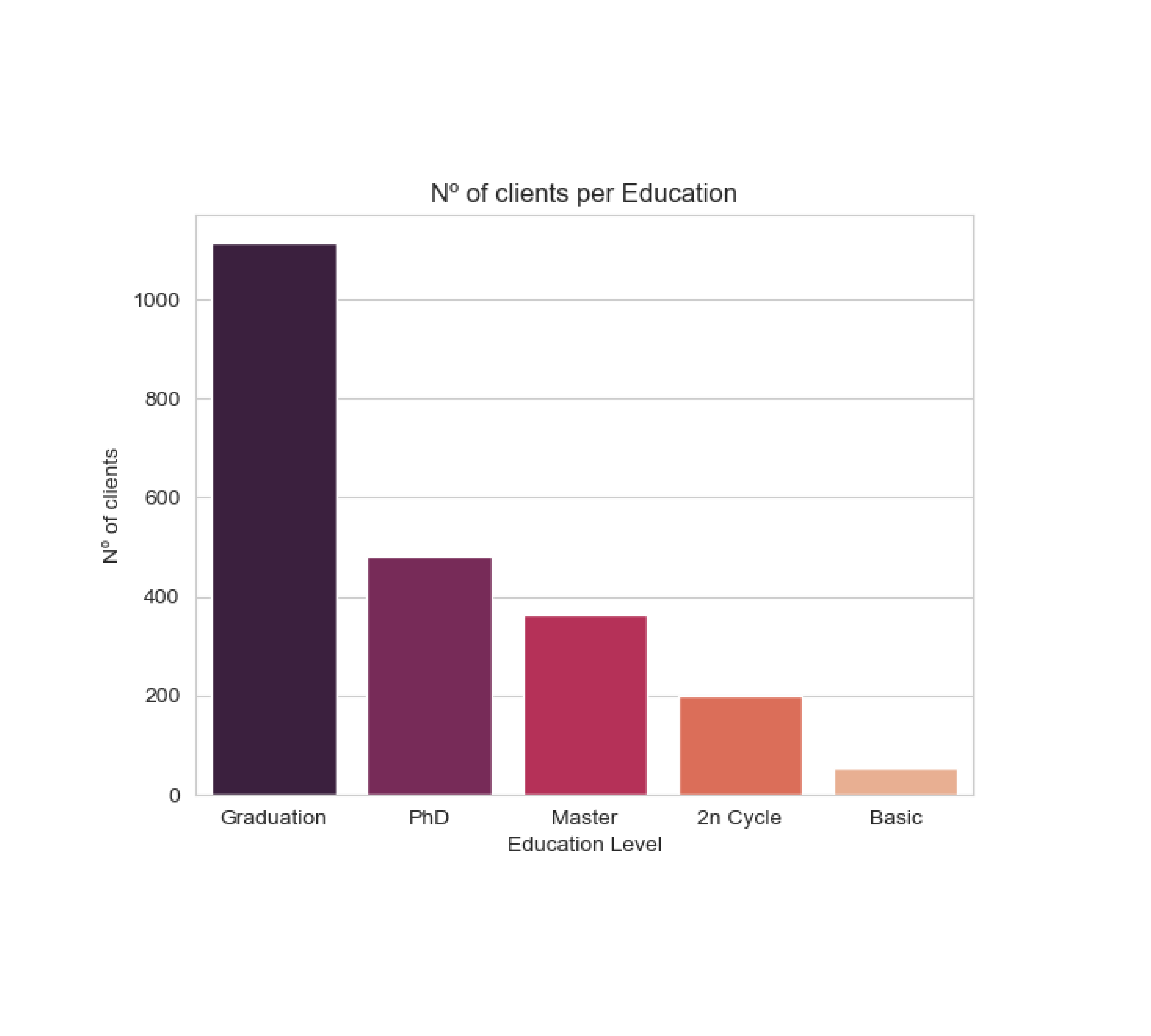
```
In [2]: income_edu = pd.read_csv("average_income_per_education.csv")

In [3]: income_mari = pd.read_csv("average_income_per_marital_status.csv")

In [4]: cmp = pd.read_csv("campaigns_statistics.csv")

In [5]: deal = pd.read_csv("deal_statistics.csv")

In [6]: pur = pd.read_csv("purchases_statistics.csv")

In [7]: sp_pur_edu = pd.read_csv("spent_amount_and_purchases_per_education.csv")

In [8]: sp_pur_mari = pd.read_csv("spent_amount_and_purchases_per_marital_status.csv")

In [9]: sp = pd.read_csv("spent_amount_statistics.csv")

In [10]: cust = pd.read_csv("type_of_customers.csv")
```

Estadísticas por educación

	education	n_of_clients	average_age	average_income	average_kids_at_home	average_teens_at_home
0	Graduation	1116	54	52720.4	0.44	0.49
1	PhD	481	58	56145.3	0.40	0.60
2	Master	365	57	52917.5	0.46	0.53
3	2n Cycle	200	52	47633.2	0.48	0.41
4	Basic	54	47	20306.3	0.63	0.09

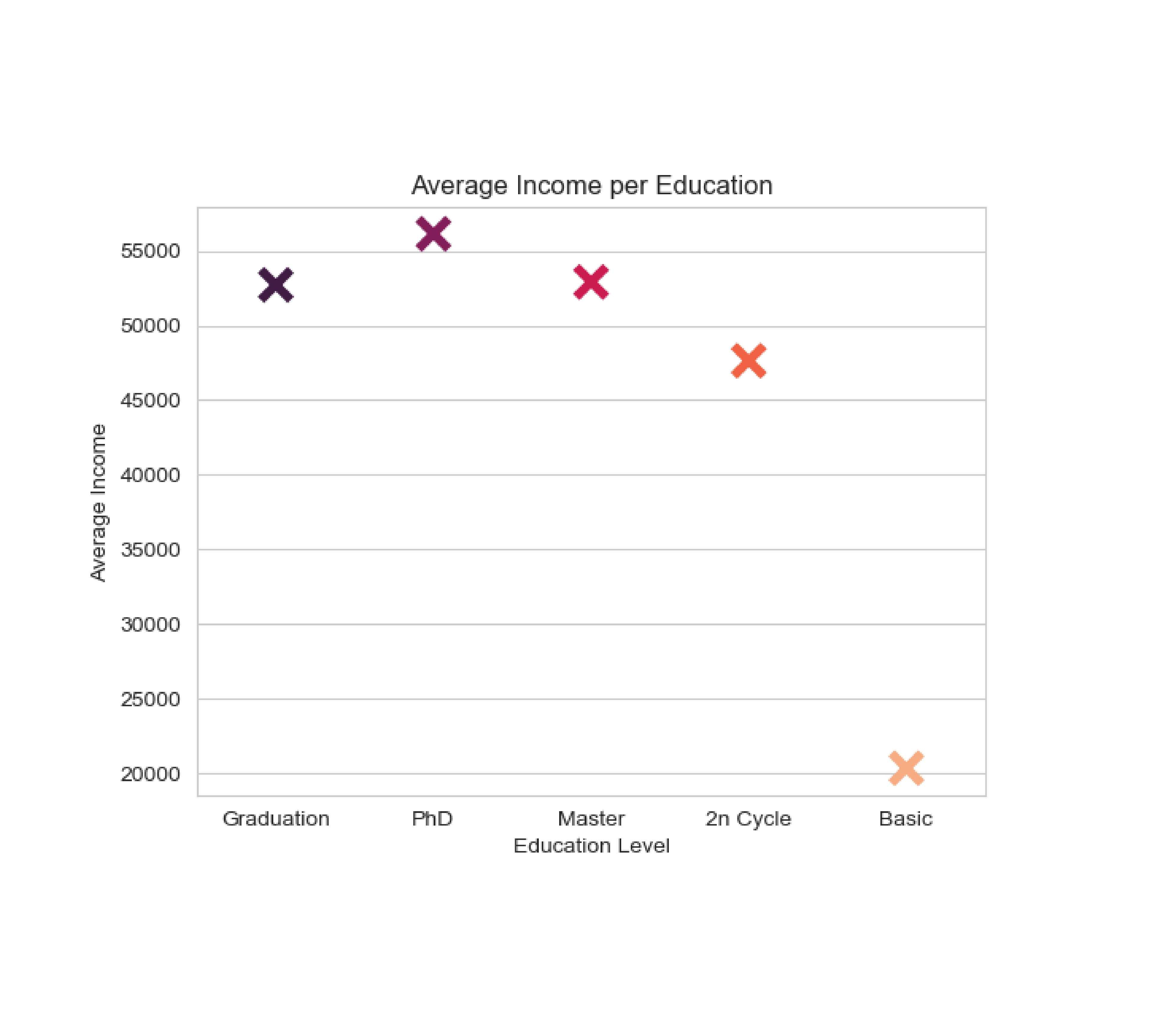
Estadísticas por educación

Número de clientes por Educación



Estadísticas por educación

Media de ingresos por Educación

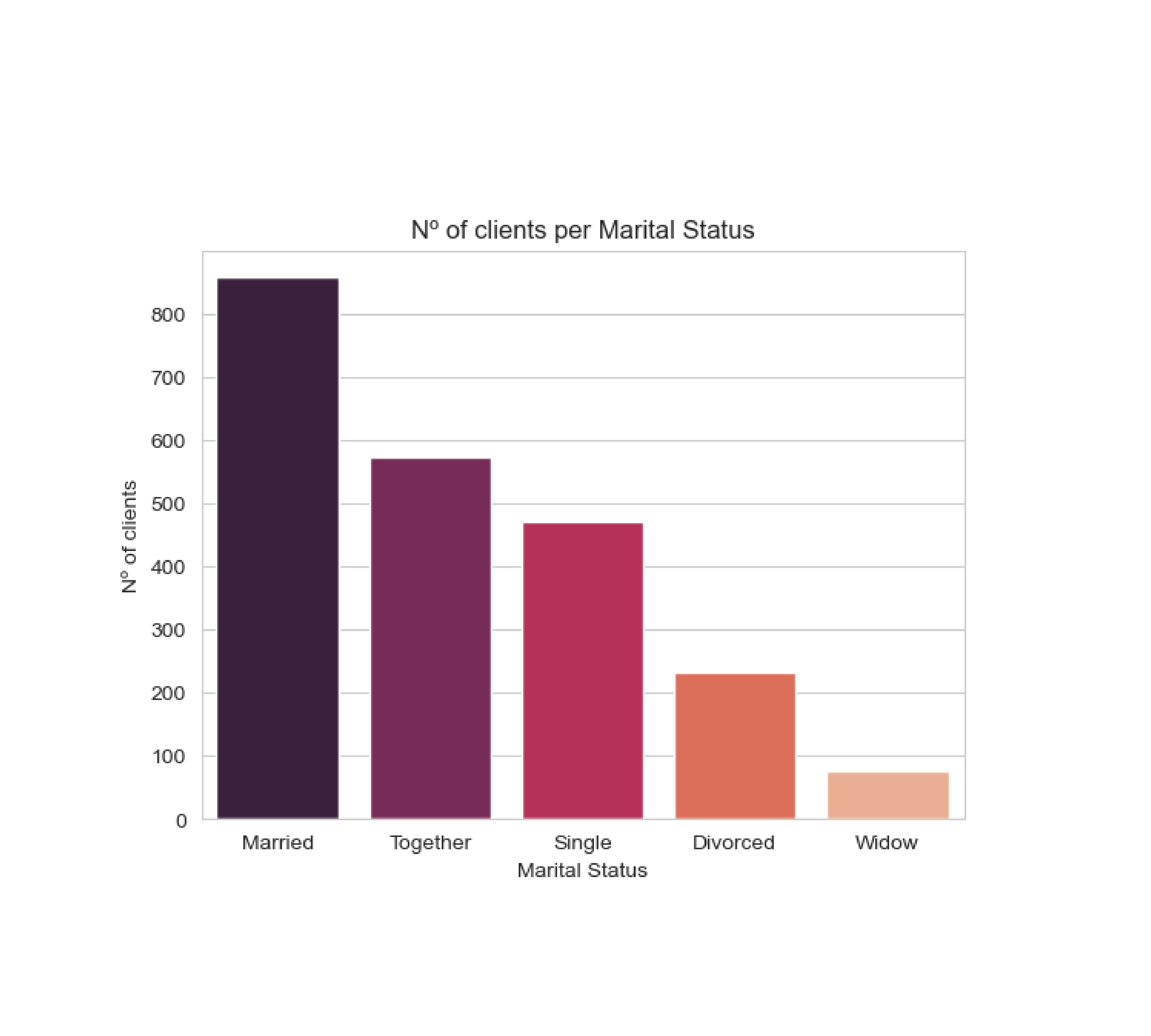


Estadísticas por estado civil

	marital_status	n_of_clients	average_age	average_income	average_kids_at_home	teens_at_home
0	Married	857	54	51725.0	0.45	0.51
1	Together	573	56	53245.5	0.45	0.53
2	Single	471	52	50995.4	0.46	0.40
3	Divorced	232	58	52834.2	0.41	0.59
4	Widow	76	65	56481.6	0.24	0.64

Estadísticas por estado civil

Número de clientes por Estado Civil

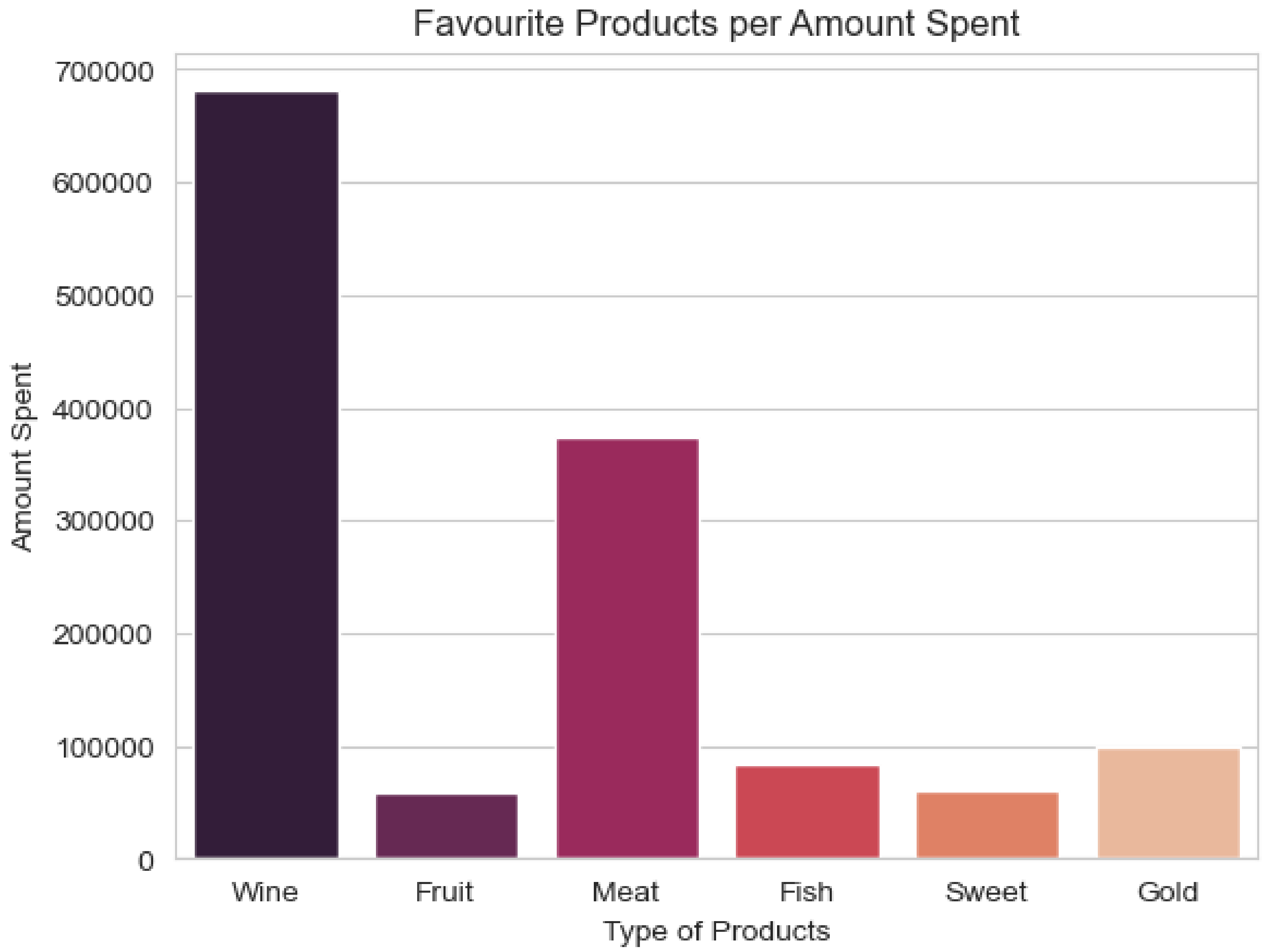


Estadísticas por estado civil

Media de ingresos por Estado Civil

Estadísticas por Gasto

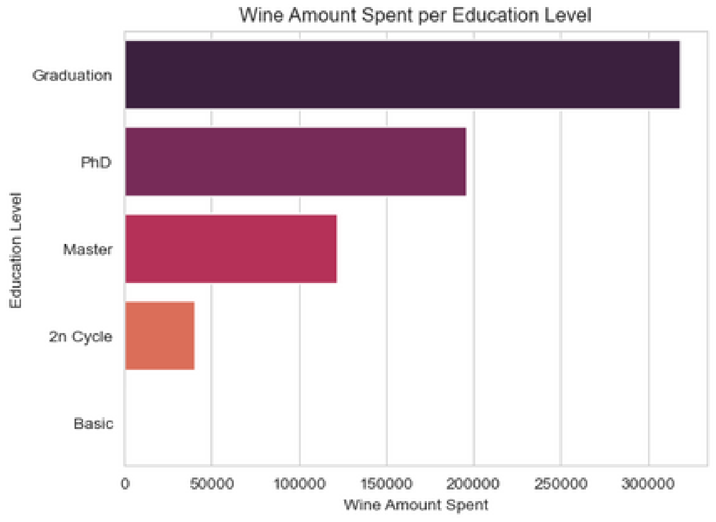
	total_wine	total_fruit	total_meat	total_fish	total_sweet	total_gold	total_all_products
0	680816	58917	373968	84057	60621	98609	1356988

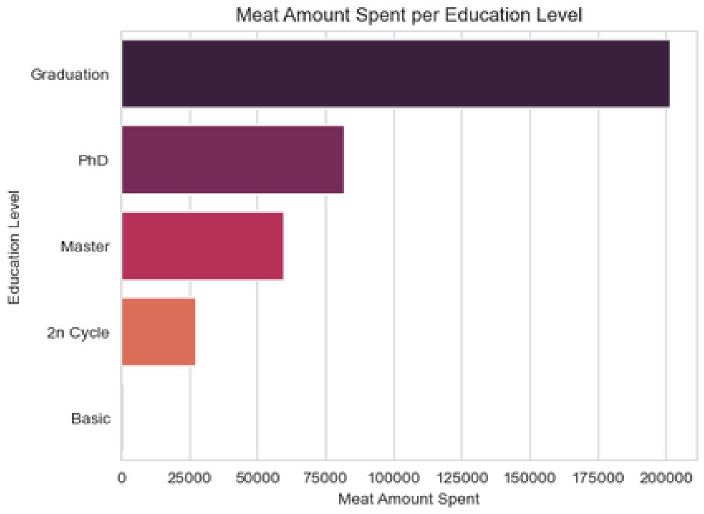


Estadísticas por Gasto

Gasto por Educación

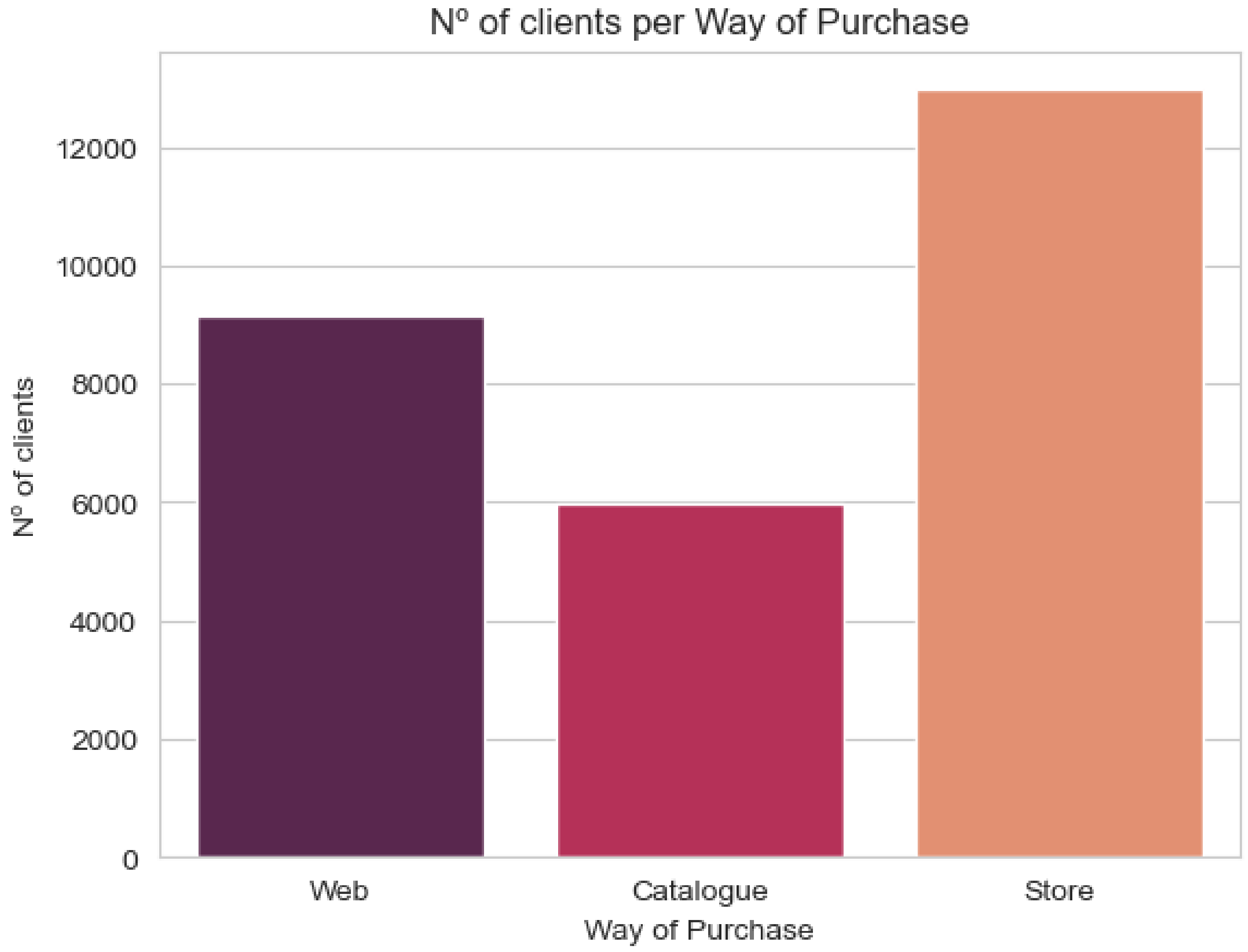
	education	total_wine	total_fruit	total_meat	total_fish	total_sweet
0	Graduation	318111	34441	201319	48453	34915
1	PhD	195874	9690	81644	12928	9787
2	Master	121538	7802	59466	11495	7595
3	2n Cycle	40169	5872	27016	9608	6945
4	Basic	391	600	618	921	654





Estadísticas por Modo Compra

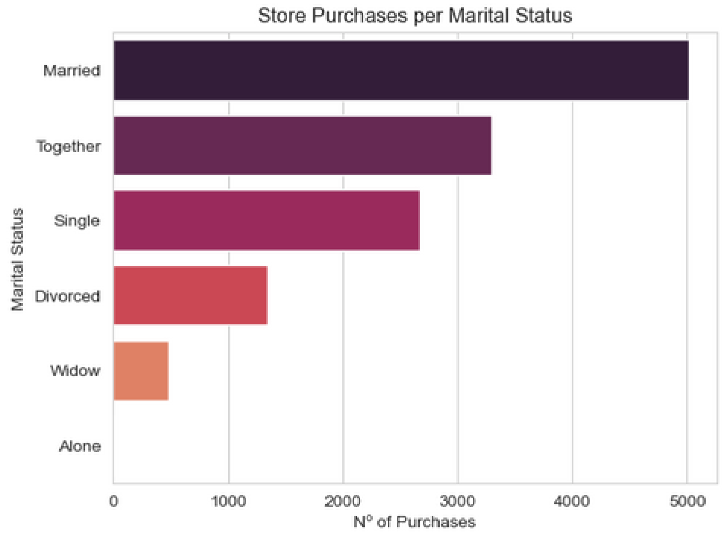
	web_purchases	catalog_purchases	store_purchases	total_purchases
0	9150	5963	12970	28083

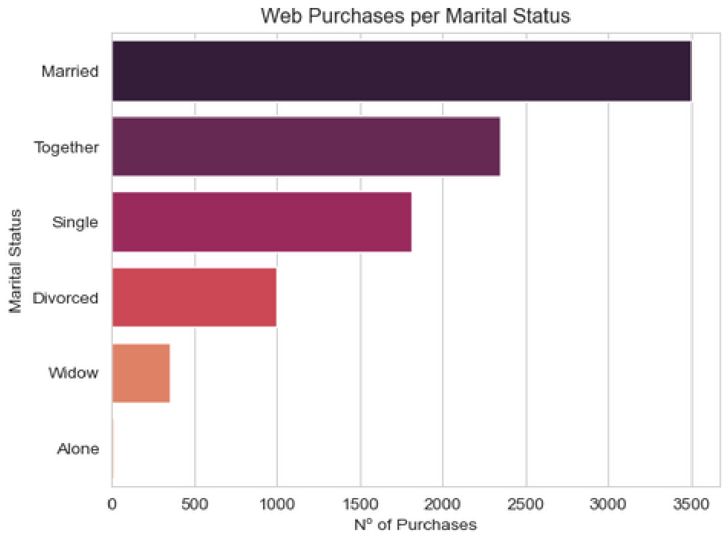


Estadísticas por Modo de compra

Compra por Estado Civil

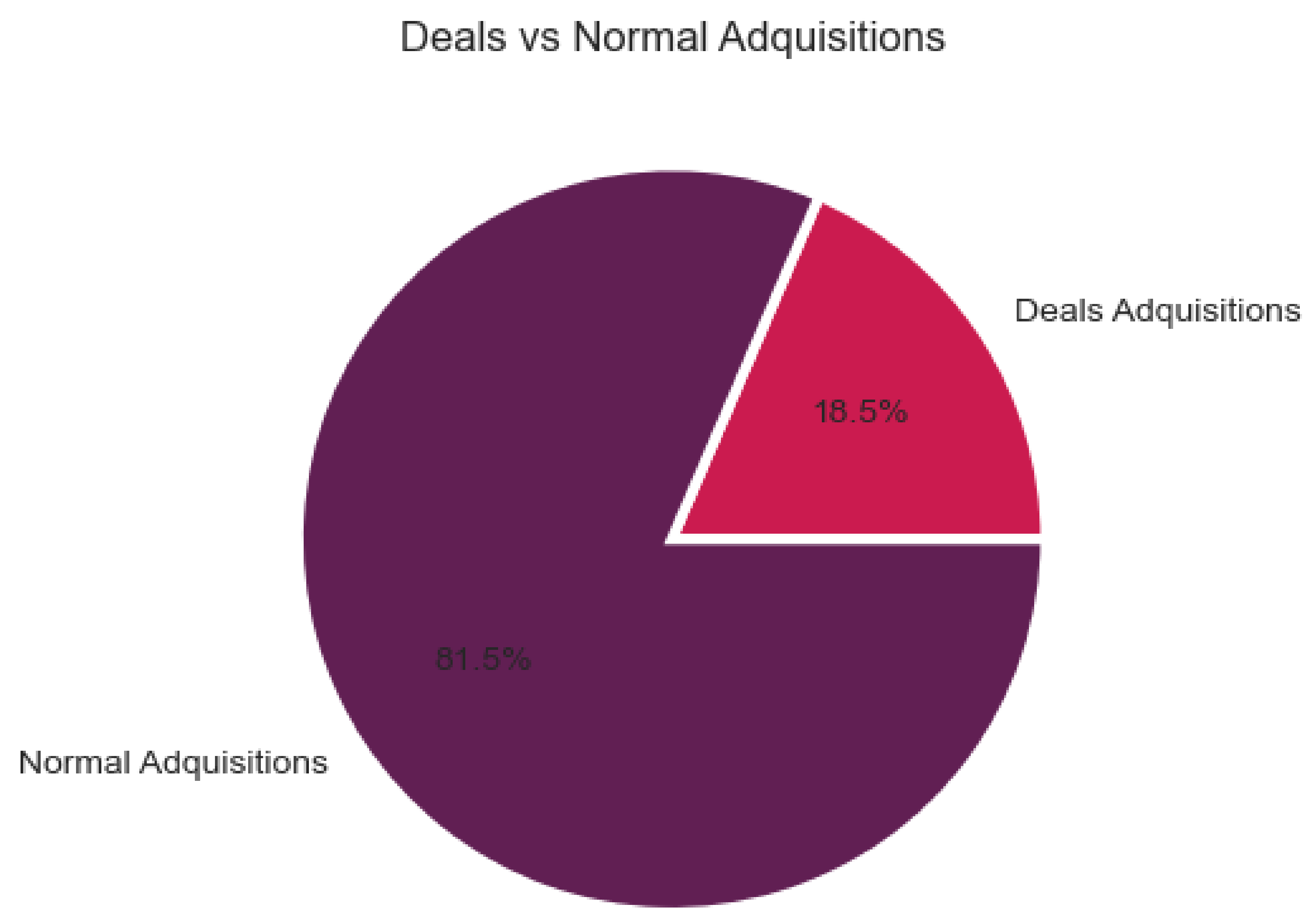
	marital_status	web_purchases	catalog_purchases	store_purchases
0	Married	3501	2254	5013
1	Together	2351	1535	3298
2	Single	1814	1240	2674
3	Divorced	1000	620	1350
4	Widow	351	251	483
5	Alone	15	2	12





Estadísticas de Oferta de Compra

	n_of_deals	total_purchases	deals_vs_purchases
0	5208	28083	18.55



Estadísticas de Aceptación de Campaña

	total_cmp1	total_cmp2	total_cmp3	total_cmp4	total_cmp5
0	144	30	163	167	163