



# ANALYZING U.S. MARKET CONSUMER SPENDING PATTERNS (2010S)





# THIS PROJECT AIMS TO EXPLORE HOW U.S. CONSUMER SPENDING EVOLVED DURING THE 2010S.

*We will see in the presentation the evolution of the transaction volume and amount during the decade and as well the market segmentation.*

*To perform the analysis, and due the volume of data +13 million rows on the transaction table, I choosed Jupyter Notebook with the **PySpark** library.*

*I used the **Extract – Transform – Load process** by connecting to the database (kaggle.com which contains millions available cases).*

*Then during the **Exploratory Data Analysis**, cleaned and transformed data into the right format **to prepare for analysis** to focus on the analysis objective.*

*As the table ,Merchant Category Codes' contains 109 unique codes, I grouped some categories to get the proper DataFrame (for technical reason – minimizing server resources, and for analytical reasons – optimizing the readiness)*

*The next step was **the data visualization in order to discover clusters, trends and more.***

*To do this, I have **created line plots, scatter plots, heatmap and a segmentation market table.***

*The the last step is **my conclusion** to answer the aim of the analysis.*



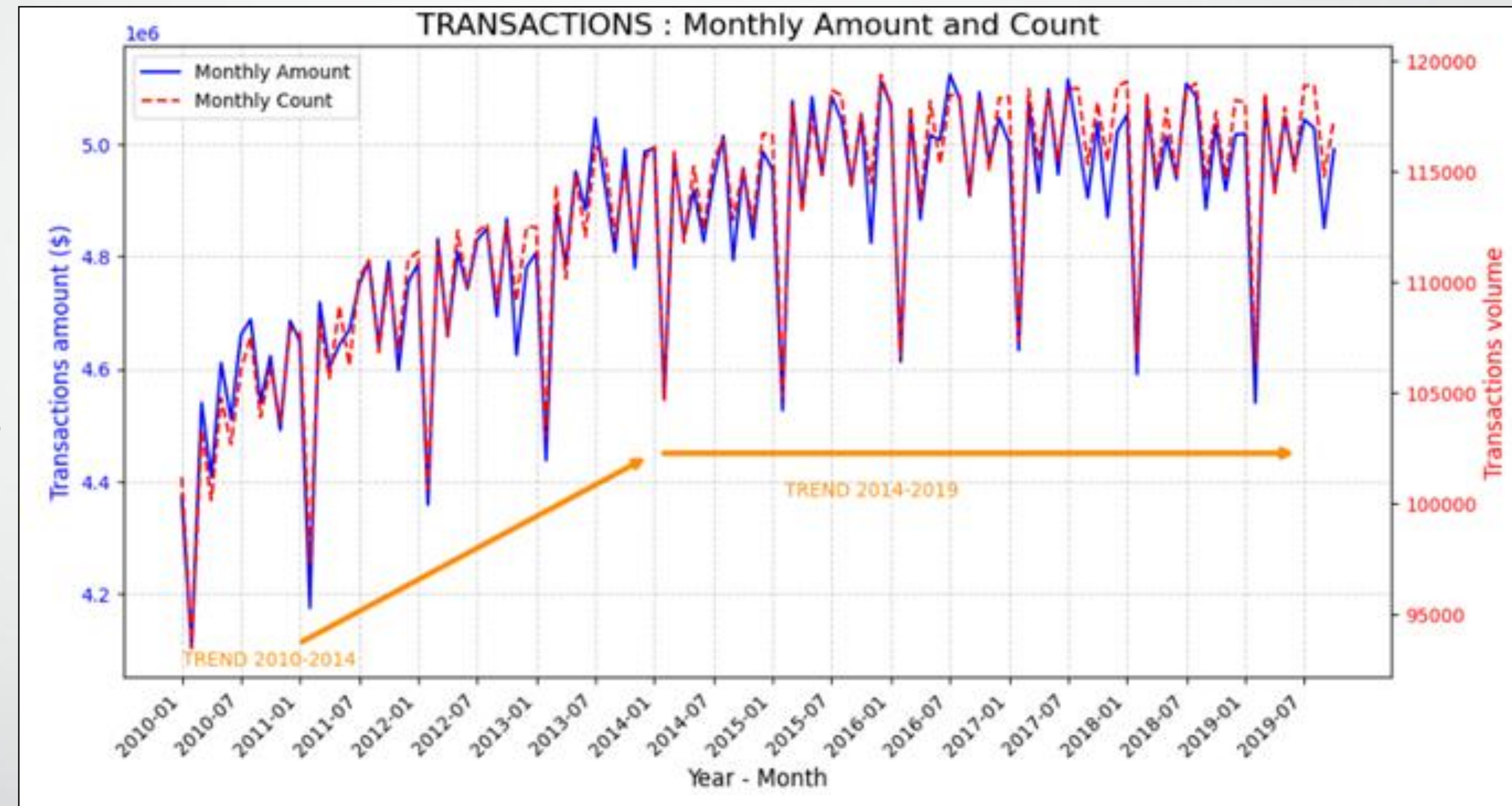
## TRENDS

*Due to a big data volume in the transaction table, i used **PySpark**, **Pandas** and **Matplotlib** libraries for this analysis*

*Observing the transaction amount exhibits a consistent trend since 2014; I included the transaction volume to confirm whether the change in trend is attributed to inflation.*

*The amounts and volume of transactions remain consistent, indicating that inflation is not the reason. Certain elements that might clarify the stabilization include:*

*Normalization of the monetary policy framework by the FED (Federal Reserve System)  
Rise in the key rate since 2015 to dampen investments.*



*During the 2010/s decade, 2 main trends appears :*

- **Growth period : 2010 – 2014**
- **Stable period : 2014 – 2019**



# CLUSTERS

To discover cluster, i used *scatter plots* using *Pandas* and *Seaborn* libraries.

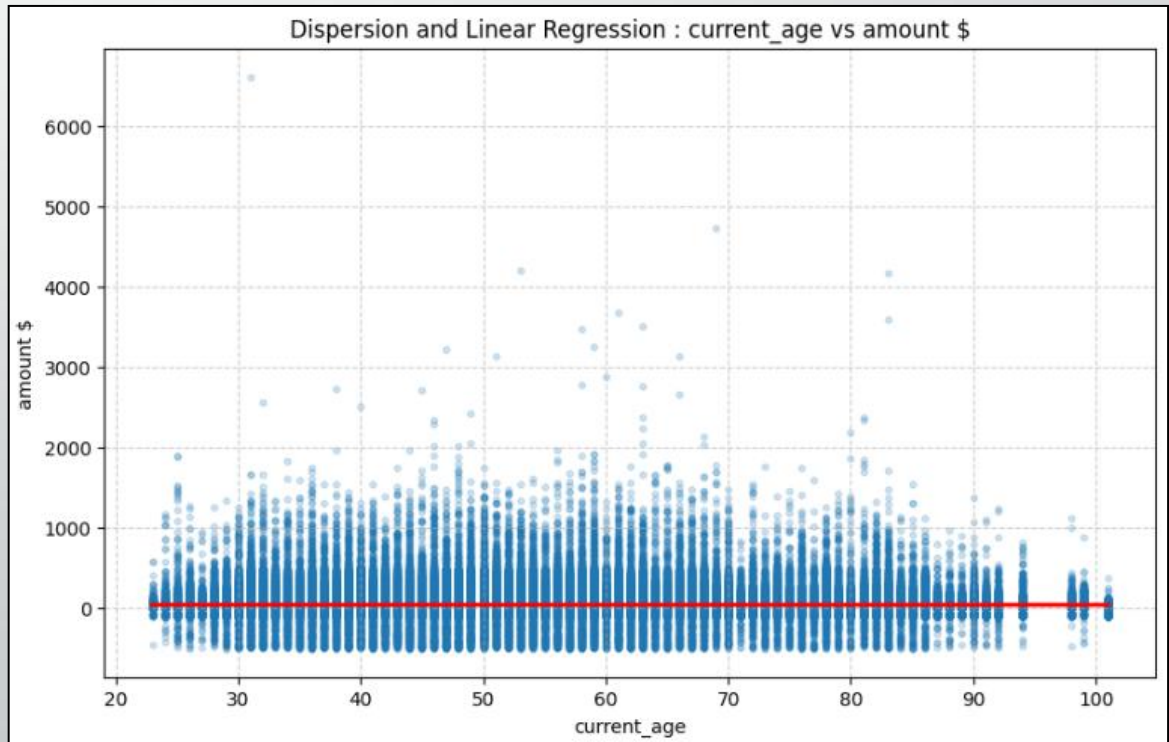
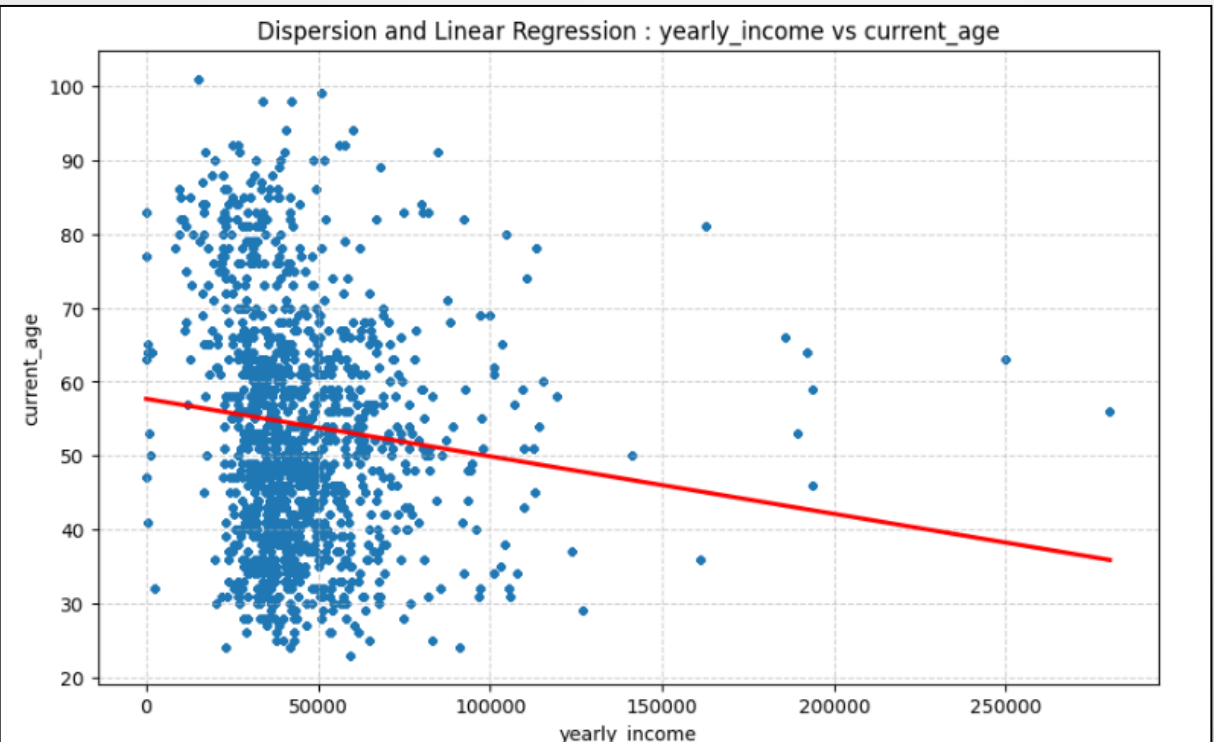
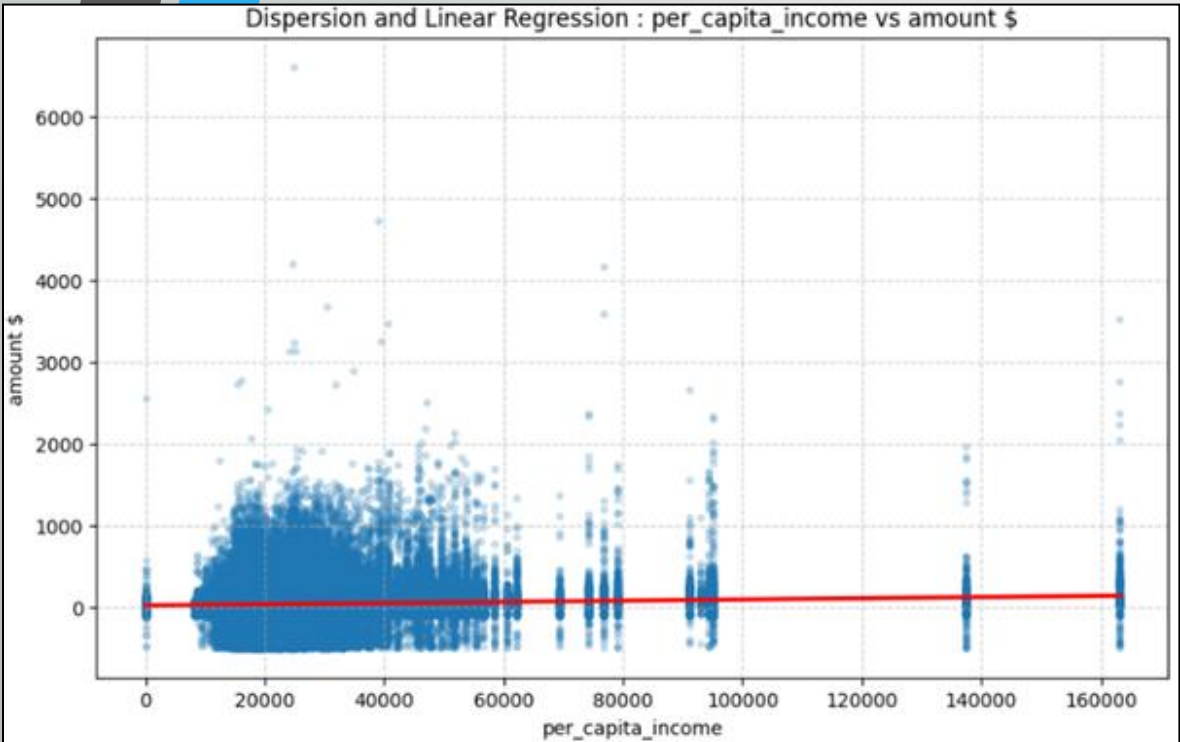
3 clusters shows a yearly income per person

- 10K\$ to 35K\$
- 35K\$ to 55K\$
- Above 55K\$

The graph typically displays an annual income ranging

- 25K\$ to 50K\$
- 50K\$ to 75K\$
- Above 75K\$

Globally a transaction amount is up to 1k\$







# HEATMAP

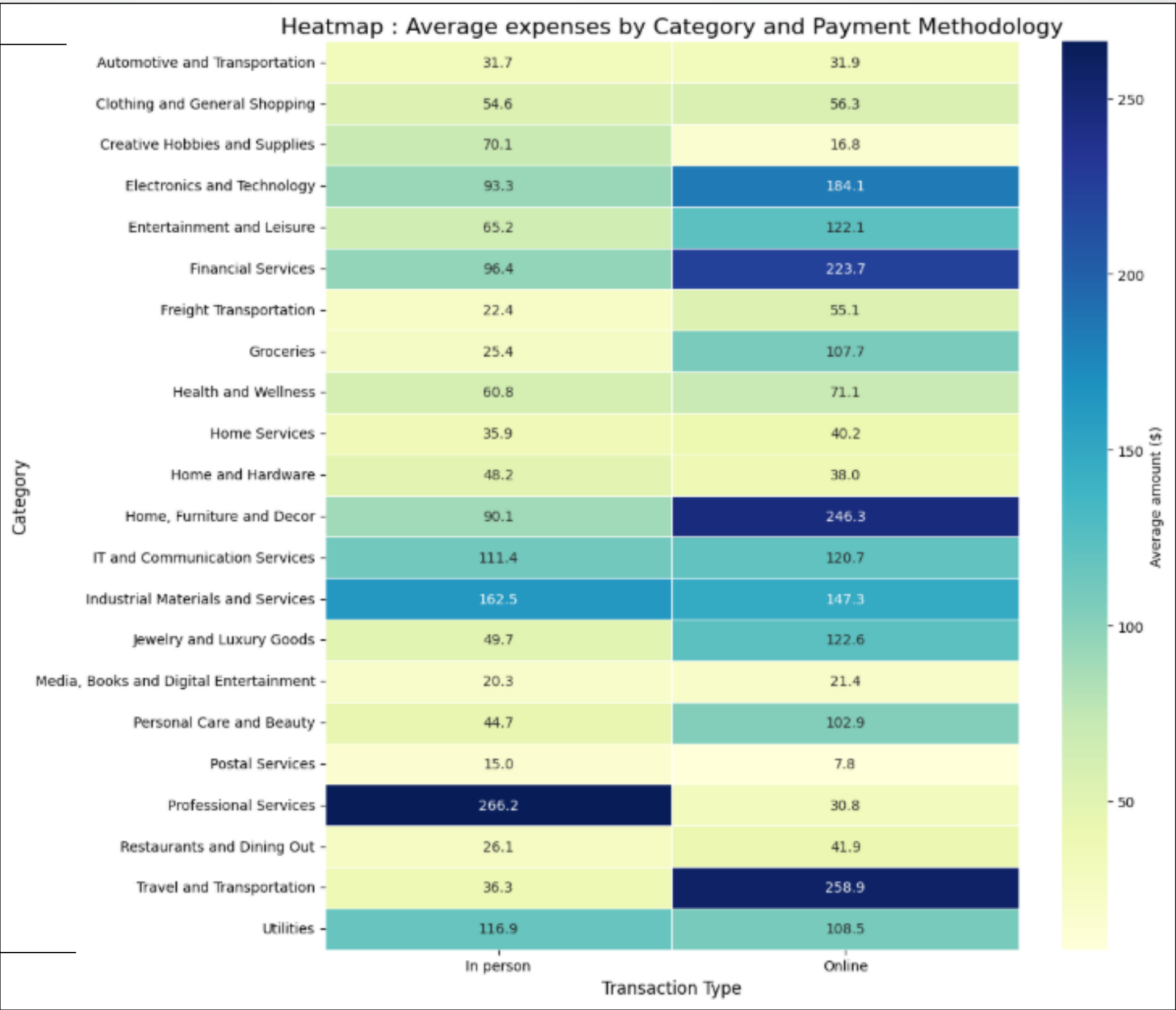
The heatmap is done using *Seaborn* library

To verify the market segmentation, I have created a heatmap which shows the Average expenses by Category and split on the methodology payment (In person / On place and Online).

The purpose was to verify if a trend will appear.

- 1. Professional Services are paid 'In person'.
- 2. For the categories below, transactions are mainly done Online

- > Travel and Transportation
- > Home, Furniture and Decor
- > Financial Services
- > Electronics and Technology





# GENDER & AMOUNT SEGMENT

For this data manipulation *Pandas* was used

Further the previous segmentation, I verified a there are others segments by gender.

- High potential categories / business :
  - **Travels** : category on which one people can afford to spend the highest
  - **Health / wellness / clothes / shopping** : people cares about themself and can afford to spend money on it
  - **Groceries** : with the highest volume of transaction.
- Overall , there are some distinctions based on gender:
  - **Men** : Media, Books, and Digital Entertainment
  - **Women** : Home Furniture, Decor, Clothing and Shopping

Market Segmentation - Sorted by Category (Max) and Gender):					
Rank	Category	Gender	Max Amount (\$)	Median Amount (\$)	Transaction Count
1	Travel and Transportation	Female	\$6,613.44	\$183.00	2,594
1	Travel and Transportation	Male	\$3,477.66	\$174.48	2,568
2	Media, Books and Digital Entertainment	Male	\$4,175.85	\$11.38	15,184
2	Media, Books and Digital Entertainment	Female	\$1,046.44	\$13.32	15,717
3	Groceries	Male	\$3,599.30	\$12.42	183,853
3	Groceries	Female	\$3,255.85	\$12.18	195,378
4	Industrial Materials and Services	Female	\$3,516.73	\$179.93	6,286
4	Industrial Materials and Services	Male	\$2,507.17	\$180.01	6,002
5	Home, Furniture and Decor	Female	\$2,378.21	\$47.93	5,289
5	Home, Furniture and Decor	Male	\$2,060.20	\$47.14	4,741
6	Clothing and General Shopping	Female	\$2,376.52	\$42.87	44,325
6	Clothing and General Shopping	Male	\$1,692.25	\$42.12	40,001
7	Financial Services	Female	\$2,345.04	\$100.00	31,158
7	Financial Services	Male	\$1,530.16	\$80.00	32,790
8	Health and Wellness	Female	\$2,340.51	\$37.14	46,217
8	Health and Wellness	Male	\$1,590.10	\$37.58	42,396
9	Professional Services	Female	\$1,893.02	\$131.90	1,389
9	Professional Services	Male	\$1,664.03	\$50.56	1,321
10	Creative Hobbies and Supplies	Male	\$1,761.58	\$39.51	3,296
10	Creative Hobbies and Supplies	Female	\$1,718.15	\$40.92	3,771



# CONCLUSION

## HOW DID THE US MARKET EVOLVE DURING THE DECADE OF THE 2010S ?

- **TREND** : 2010 – 2014, the market trend of transaction amount was increasing and from 2014 until 2019, the trend is stable
- The **standard amount** spent for transactions is up to 1000 dollars / 1500 dollars

## WHAT ARE THE MARKET SEGMENTS ?

- People are able to spend more on **travelling and health / wellness / clothes / shopping**
- Professional Services are paid 'in person' while the categories below are mainly done **online**:
  - Travel and Transportation
  - Home, Furniture and Decor
  - Financial Services
  - Electronics and Technology
- **Groceries is the category with the highest volume of transactions.**
- Depending on the category and **gender**, some **trends are visibles**.
- Then I will create an interactive dashboard in Power BI to share this conclusion.
- More inquiries arise following the examination:
  - What has contributed to the stability of the US market trend since 2014?
  - To address this new question, an alternative analysis can be carried out by examining the annual income progression by client\_id to evaluate purchasing





# THANK YOU

