**29 Group**

**Participants:**

**Luisa Maria Carabali Sanchez**

**Alejandro Camargo García**

**Cristian Camilo Sarmiento**

**Hector Julio Melo Ocampo**

**Ramon Manuel Sandoval**

**Norma Ximena Rios Cotazo**

**Topic: Consumer behavior on E-Commerce**

**Research question**

What aspects of consumers’ online behavior is useful for businesses to better understand their customers and predict consumer trends, spending habits, variables related to the shopping?

Objectives:

* Data understanding and cleaning
* Analysis and modeling
* Build prediction models
* Build dashboards to visualize the insight

Datasets to use:

**Online Costumer Behavior**

**Some Data overview:**

**Events**

Columns: Index(['timestamp', 'visitorid', 'event', 'itemid', 'transactionid'], dtype='object')

Events Size: 2.756.101

Events Unique visitorid: 1.407.580

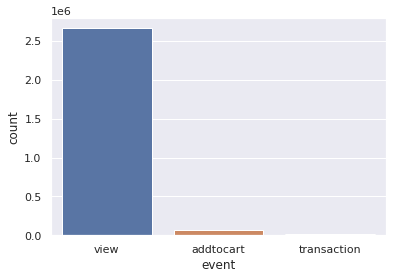
Averrage events per visitor: 2.0

Events type: ['view' 'addtocart' 'transaction']

view 2.664.312

addtocart 69.332

transaction 22.457



Nulls

timestamp False

visitorid False

event False

itemid False

transactionid True

**Category Tree:**

Columns: Index(['categoryid', 'parentid'], dtype='object')

Events Size: 1669

Unique category: 1669

**Item Properties**

Index(['timestamp', 'itemid', 'property', 'value'], dtype='object')Item Properties:

Columns: Index(['categoryid', 'parentid'], dtype='object')

Events Size: 1669

All numerical values were marked with "n" char at the beginning, and have 3 digits precision after decimal point, e.g., "5" will become "n5.000", "-3.67584" will become "n-3.675". All words in text values were normalized and hashed

**Brazilian ecommerce**

**Customer**

Customer size: 99.441

Columns: Index(['customer\_id', 'customer\_unique\_id', 'customer\_zip\_code\_prefix',

'customer\_city', 'customer\_state'],

dtype='object')

Customer State: ['SP' 'SC' 'MG' 'PR' 'RJ' 'RS' 'PA' 'GO' 'ES' 'BA' 'MA' 'MS' 'CE' 'DF'

'RN' 'PE' 'MT' 'AM' 'AP' 'AL' 'RO' 'PB' 'TO' 'PI' 'AC' 'SE' 'RR']

**Entity relationship diagram:**

