MELI
Clustering

Carolina Jiménez Moreno

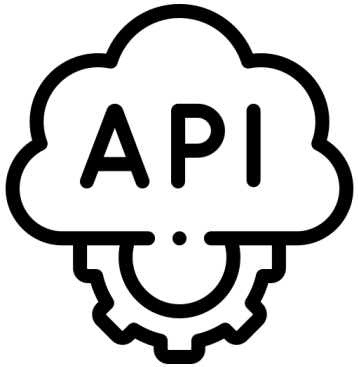
Objective

01 Searching for and grouping these elements can improve the user experience.

02 There are similar or identical products on the market sold by different vendors

Data Collection

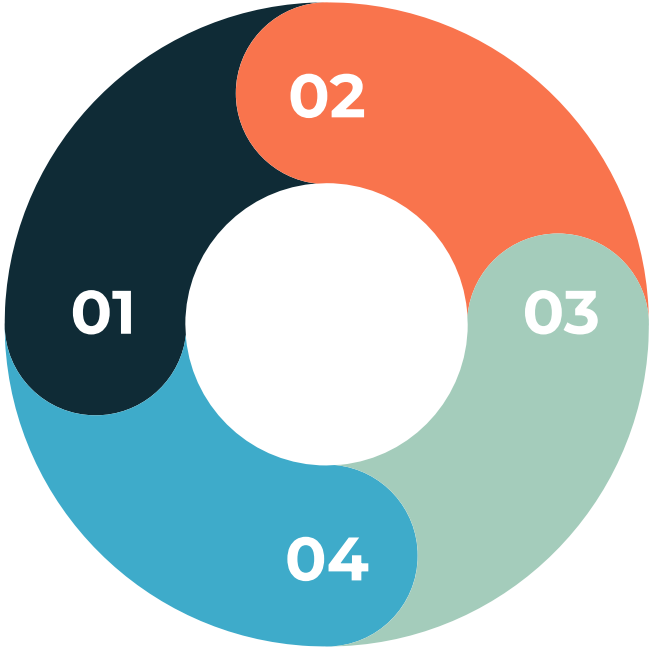
- Accessed the MercadoLibre API to collect data about items, categories, and item attributes.



Exploratory Data Analysis (EDA)

Examined the basic structure of the dataset.

Summarized its statistics.

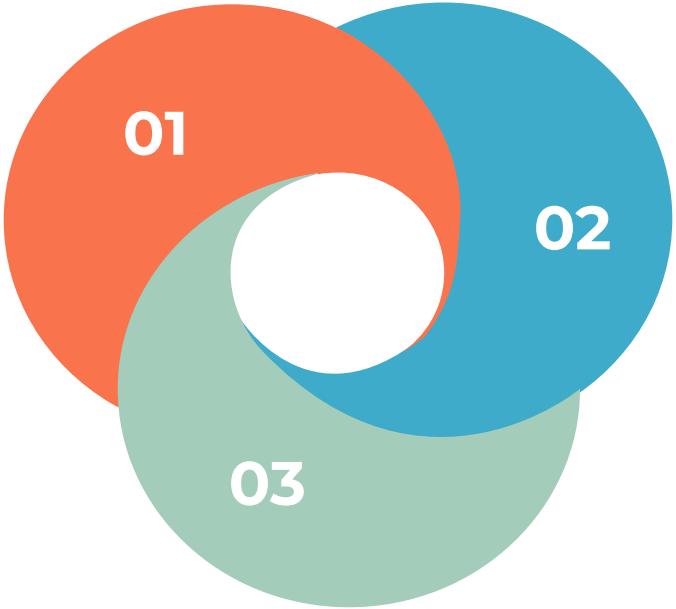


Identified missing values.

Visualized the distributions and relationships between different variables.

Data Preprocessing

Load the dataset into a Pandas DataFrame



Elimination of irrelevant columns

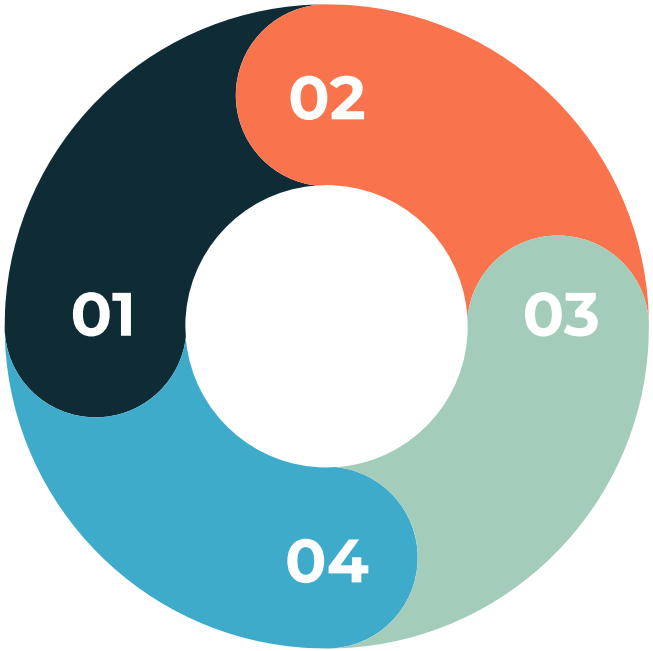
Handling null values by replacing them with empty strings

Feature Extraction Procedure

- This section outlines the procedures followed for feature extraction for both text and image data.

Exploratory Data Analysis (EDA)

Concatenation of Text Variables: ('title', 'condition', 'listing_type_id', 'buying_mode', and 'domain_id') into a single column named 'text_features'.



Text preprocessing:
* Conversion of words to lowercase
* Removal of leading and trailing whitespaces
* Removal of punctuation marks
* Remove stopwords
* Expansion of contractions
* Remove of special characters.

Tokenization and Lemmatization

Obtaining Text Features

Image Feature Extraction

Loading Pre-trained CNN Model: VGG16 pre-trained Convolutional Neural Network (CNN) model for extracting image features.

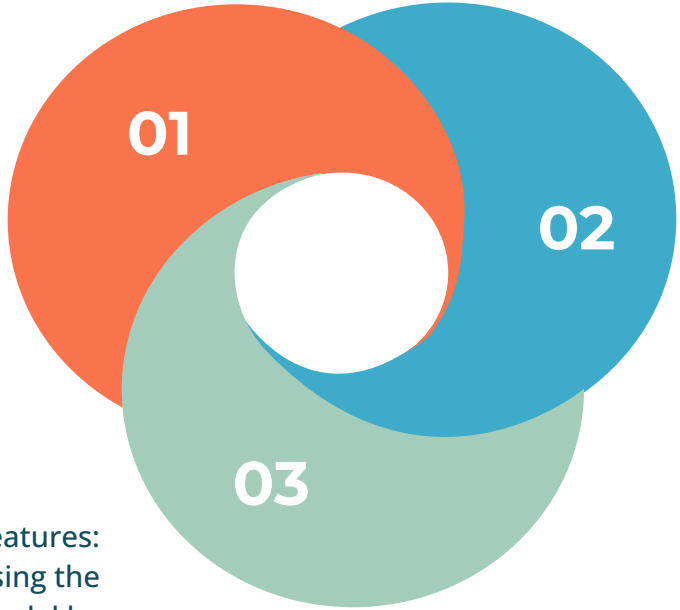
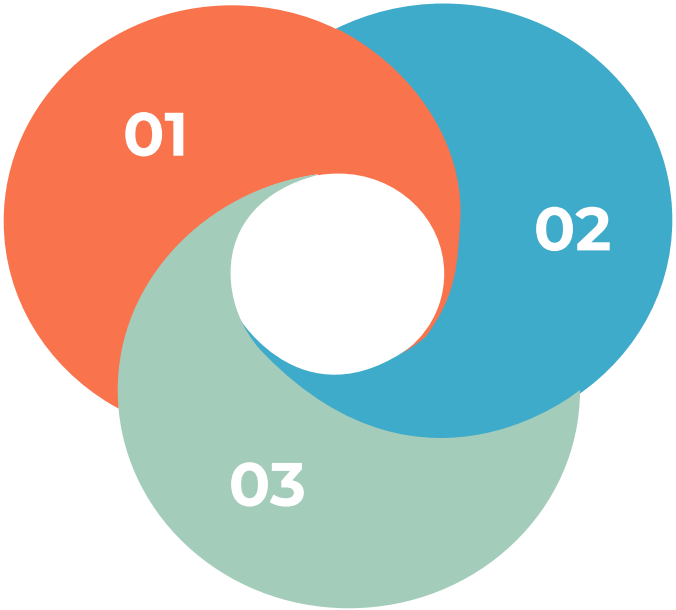


Image Preprocessing: Resized the images to 224x224 pixels and normalized pixel values.

Extracting Image Features: Extracted features using the pre-trained VGG16 model by passing the preprocessed image through the CNN model.

Clustering

Applying clustering algorithms: K-Means



Determine the optimal number of clusters

Evaluate the cohesion and separation of clusters

Evaluation

- Evaluate the quality of the groups formed using metrics such as purity, homogeneity and completeness

Thank you. Please feel free to ask any questions.