

Exploratory Data Analysis of AliExpress

*Mohammad Ghadieh✍*

Introduction: ✋

**AliExpress, a leading global e-commerce platform, offers a diverse range of products at competitive prices. With reliable international shipping and a user-friendly interface, it has become a go-to destination for online shoppers worldwide. By harnessing the power of Exploratory data analysis (EDA), AliExpress provides valuable insights into consumer behavior, market trends, and product preferences. Through this project, we aim to uncover patterns, forecast demands, and empower businesses to make data-driven decisions for enhanced customer satisfaction and growth. Explore the vast potential of AliExpress data with us as we delve into the world of online shopping and unlock actionable insights.**

Dataset Sources: <https://www.kaggle.com/datasets/abdullahbuzaid/ali-express-data>

Dataset info: 864,696 entries, 17 columns

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 id 864270 non-null int64

1 storeId 864270 non-null int64

2 storeName 864270 non-null object

3 title 864270 non-null object

4 rating 864270 non-null float64

5 lunchTime 864270 non-null object

6 category 0 non-null float64

7 postCategory 864270 non-null int64

8 sold 864270 non-null object

9 price 864270 non-null float64

10 discount 864270 non-null int64

11 shippingCost 864270 non-null object

12 imageUrl 864270 non-null object

13 storeUrl 864270 non-null object

14 category\_name 864270 non-null object

15 category\_id 864270 non-null int64

16 type 864270 non-null object

dtypes: float64(3), int64(5), object(9)

memory usage: 112.1+ MB

Note: *It appears that the majority of columns in the dataset are populated with non-null values; however, there are instances where certain entries are mess or marked as 'None.'*

Methodology:

**1.** **Data Preprocessing:** Clean and preprocess the dataset, handling missing values, removing duplicates, and standardizing data formats.

**2.** **Descriptive Analysis:** Conduct descriptive statistics to summarize the dataset's main characteristics.

**3. Exploratory Data Analysis:** Explore the dataset through visualizations, including bar charts, histograms, scatter plots, and heatmaps. Examine relationships, trends, and patterns to identify key insights.

**4. Statistical Analysis:** Apply statistical tests, such as t-tests or chi-square tests, to determine significant differences or associations between variables.

**5. Recommendations:** Based on the analysis of the AliExpress data, identify trends, customer preferences, and areas for improvement to inform strategic decision-making and enhance the overall customer experience.

Questions:

1. What are the top 10 stores with the highest ratings?

2. Which category has the highest average discount?

3. What is the average price of items in each category?

4. How many products are sold in each category?

5. What is the average shipping cost for items in each category?

6. Which stores offer the highest discounts on their products?

7. How many items are sold by each store?

8. What is the average rating of stores in each category?

9. What is the distribution of product ratings across all categories?

10. Which category has the highest number of products listed?

✌