JENNIFER SOUSA CAMERON
PORTOFOLIO

ABOUT ME

Hi, I’m Jennifer, I’m a data analyst with a background in aerospace design engineering with a passion for fashion and forecasting trends. I have a strong foundation in analyzing complex problems, visualizing solutions, and ensuring attention to detail. By leveraging my expertise in design engineering and combining it with data analytics, I offer a unique perspective to help fashion brands optimize their inventory, stay ahead of competitors, and make data-driven decisions. I thrive in cross-functional collaborations, and my ability to translate complex data into actionable insights sets me apart as a fashion-focused data analyst. I’m eager to use my expertise to analyze large data sets and give insight on personnel trends that can guide business decisions and lead to success.

SKILLS

Excel

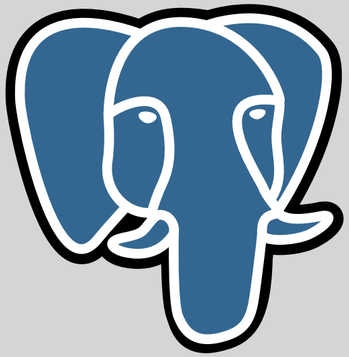
Pivot Tables, VBA

Tableau

Visualizations, dashboards

Python

Anaconda, Jupyter, Pandas, Matplotlib, Seaborn, Numpy



SQL

PostgreSQL

MY WORK

Enhancing decision-making processes through data-driven visualization reports.

Process

1. Data Preparation

The first step was to understand and define the objective, assess data requirements, and clean the data.

2. Data Analysis

Initially, I conducted Geographic Visualizations to explore geographic variables using choropleth maps. Following that, I performed regression analysis using supervised machine learning, but the results did not contribute significantly to the analysis. Next, I attempted Cluster analysis using unsupervised machine learning, but faced challenges due to the limited number of numerical columns in the dataset. Ultimately, I conducted a time-series analysis in Python and prepared time-series data for forecasting.

3. Results

Despite deriving meaningful insights from the analysis, the limitations of the dataset became evident during the machine learning sections. The stakeholders were informed about these limitations and provided with advice based on the analysis results.

Background

Context

- I analyzed a publicly available Brazilian ecommerce dataset on Kaggle comprising 100,000 orders made at the Olist Store. The dataset encompasses information from multiple marketplaces in Brazil, spanning the years 2016 to 2018.

Project Brief Link

TOOLS

Python

Tableau

SKILLS

- Conducted regression analysis in Python using supervised machine learning.
- Conducted cluster analysis using unsupervised machine learning.
- Created engaging visuals to form an interactive data dashboard.

Dataset

Github Repository

Tableau Presentation

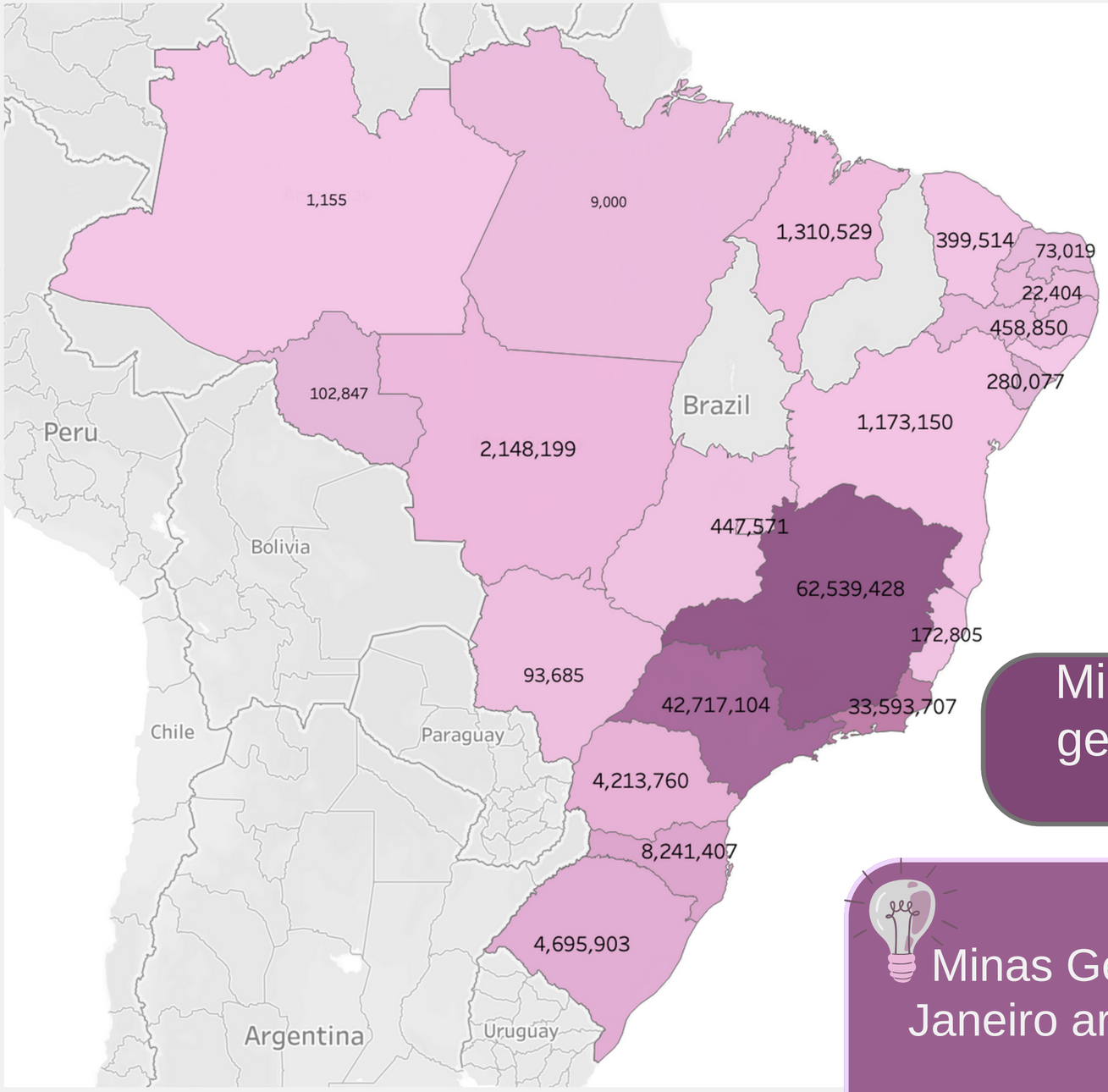
Background

Problem Statements

- Find out if seasonality played a factor in sales surges.
- Determine which states generated the highest revenue.
- Identify the top-selling product.
- Discern which product category generates the most revenue.

Data Limitations

- Limited numerical value columns
- Insufficient data is available to perform a seasonal trend and cluster analysis, as the dataset only encompasses the years 2016 to 2018
- The order column did not specify if the order had multiple items, making it challenging to understand customer behavior and buying patterns.



Hypothesis

1.States with larger population will generate a higher revenue.
2.Sales will experience a surge during the Holiday season.

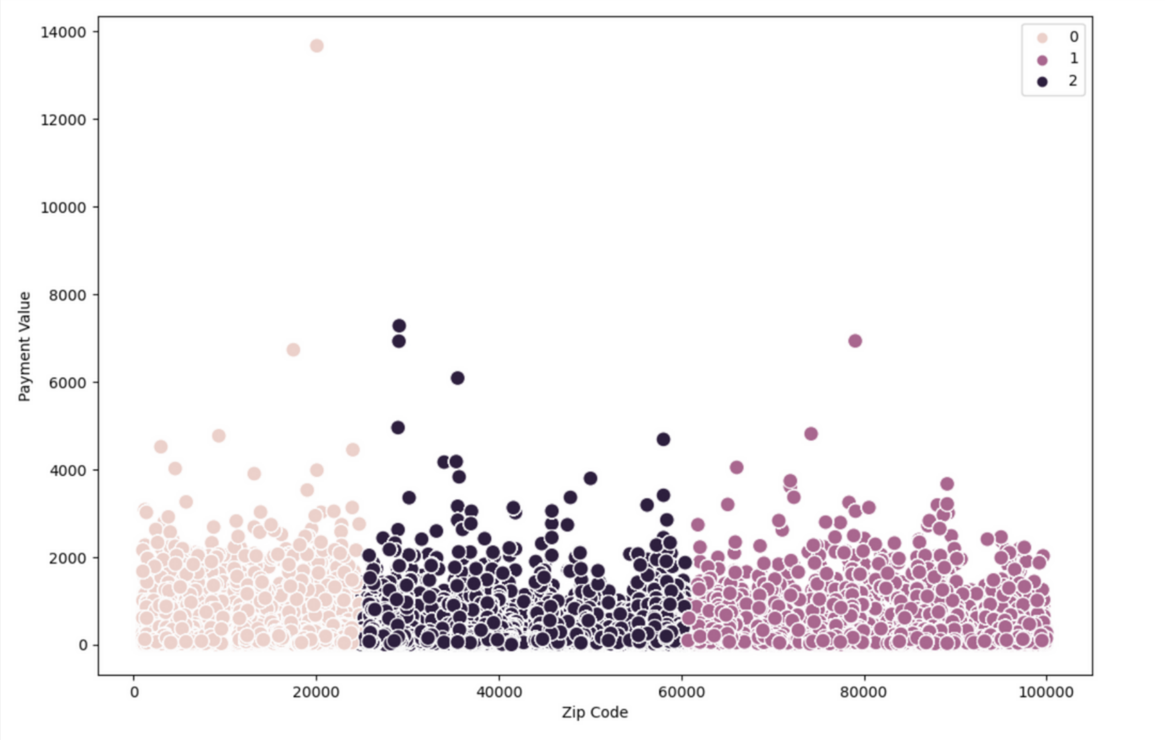
Goal

This analysis will provide insight on sales, consumer, and data trends by year and state.

Minas Gerais, Sao Paulo, and Rio de Janeiro generated the highest sales, thus proving the first hypothesis.

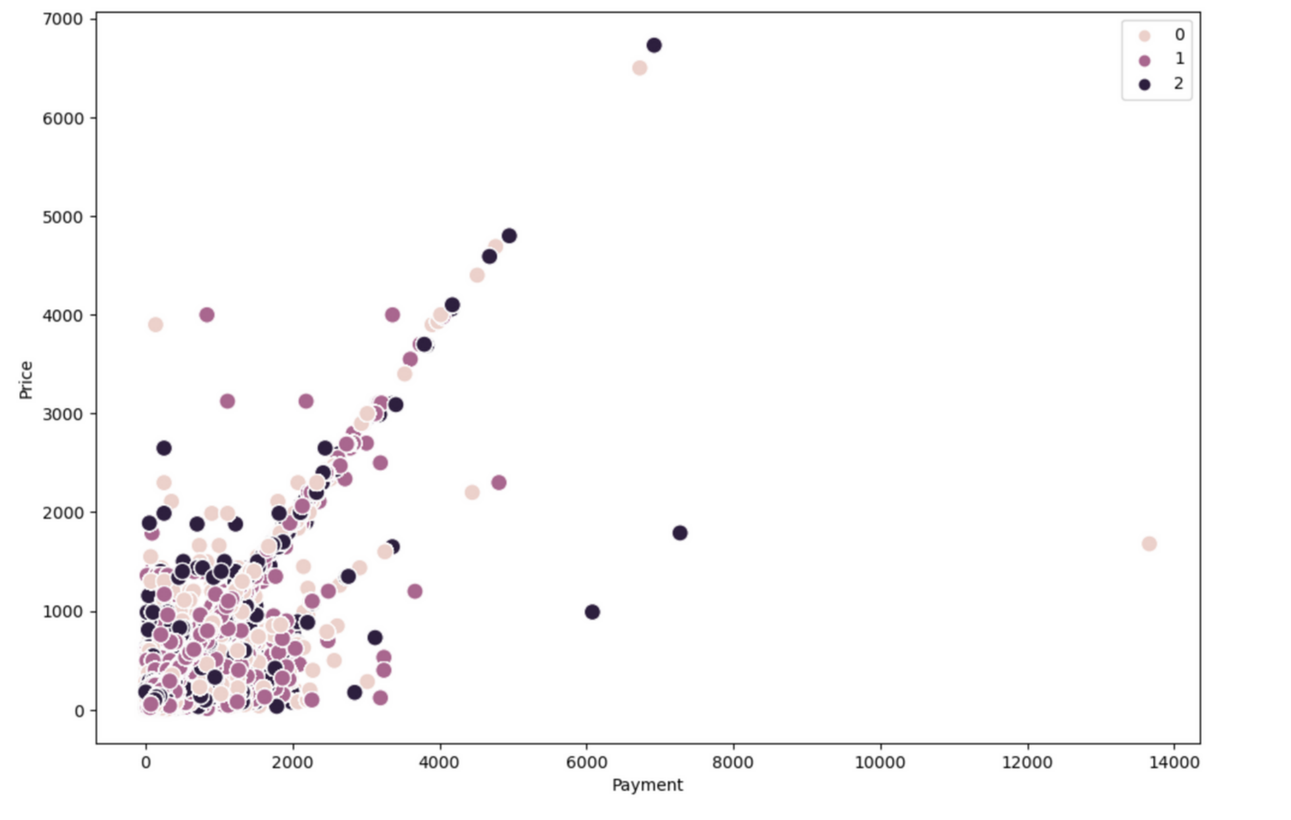
Minas Gerais, Sao Paulo, and Rio de Janeiro are the most populous states in Brazil.

Cluster Analysis

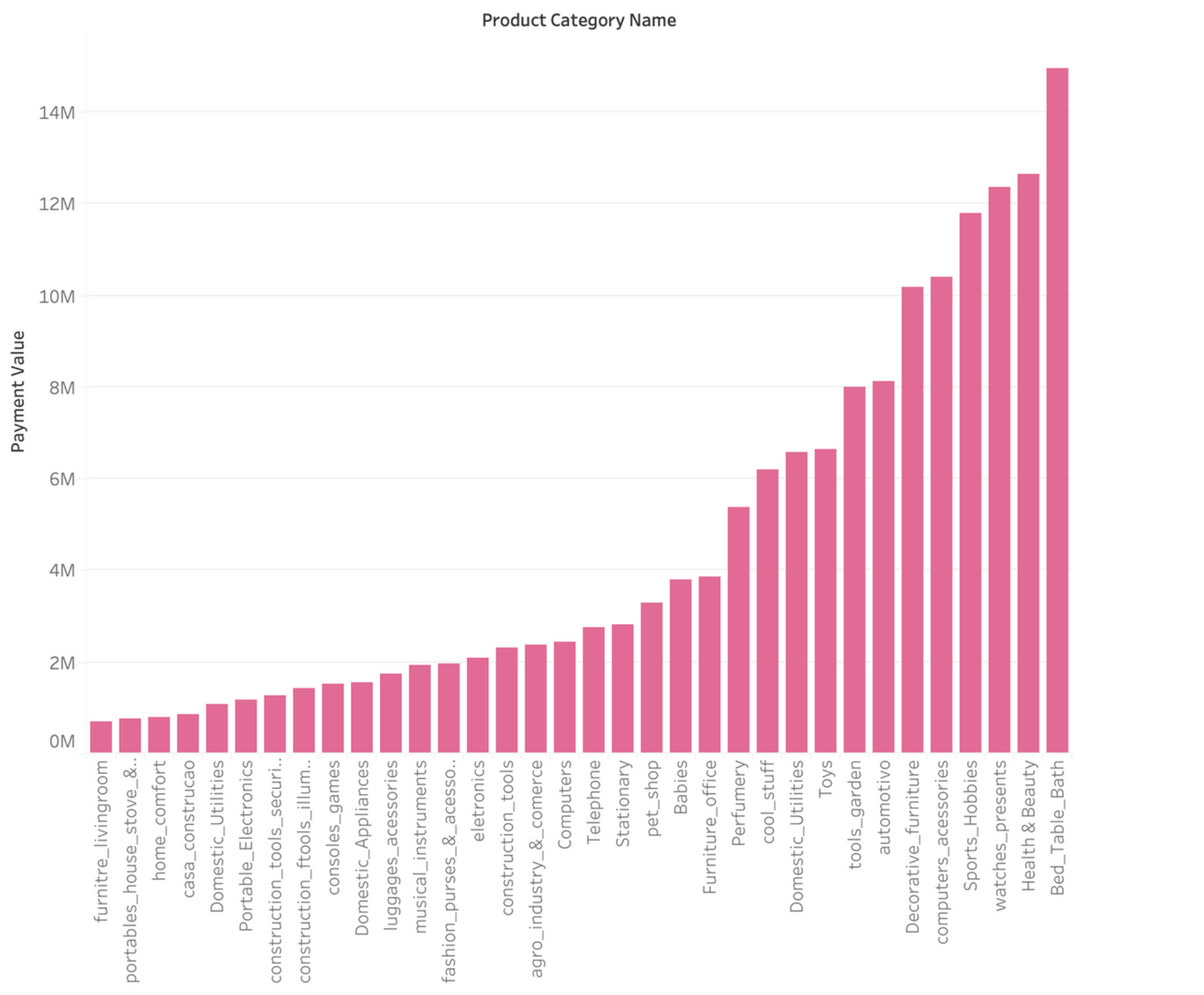


Payment Value and Zip Code were divided into three distinct categories.

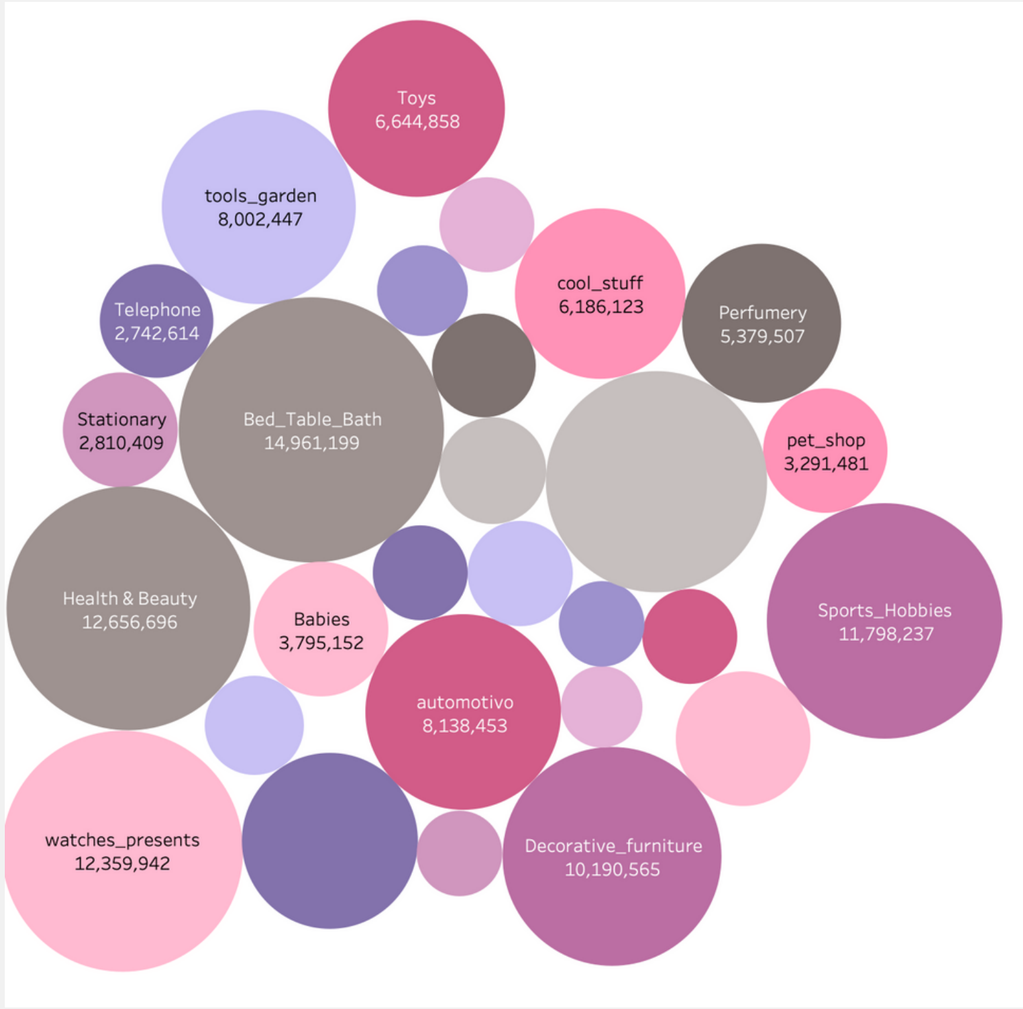
When comparing the payment value , the resulting cluster lacked clear distinctions and did not provide meaningful insight.



Analysis



The product category Bed_Table_Bath has the highest sales in the combined years 2016-2018.



Insight

Data Collection

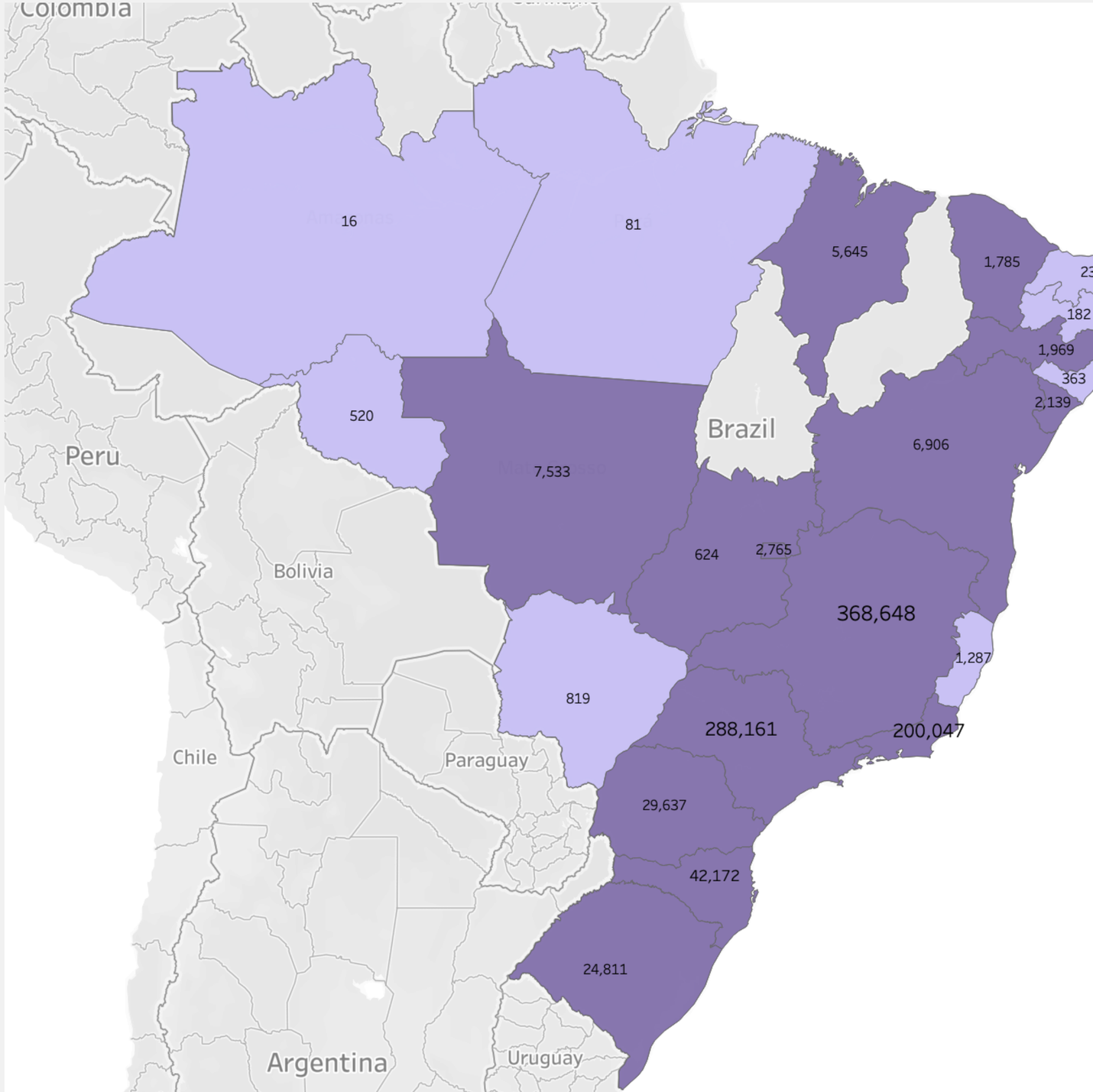
- Expand data collection and analysis beyond 2018
- Gather marketing data to evaluate marking strategy effectiveness.

Product

- The most popular category was domestic utility.
- The highest revenue was generated by the Bed_Table_Bath.

Consumer

- Minas Gerais, Sao Paulo, and Rio de Janeiro emerge as the states with the highest generated payment values.
- August was the highest revenue month.



Insight

What went well

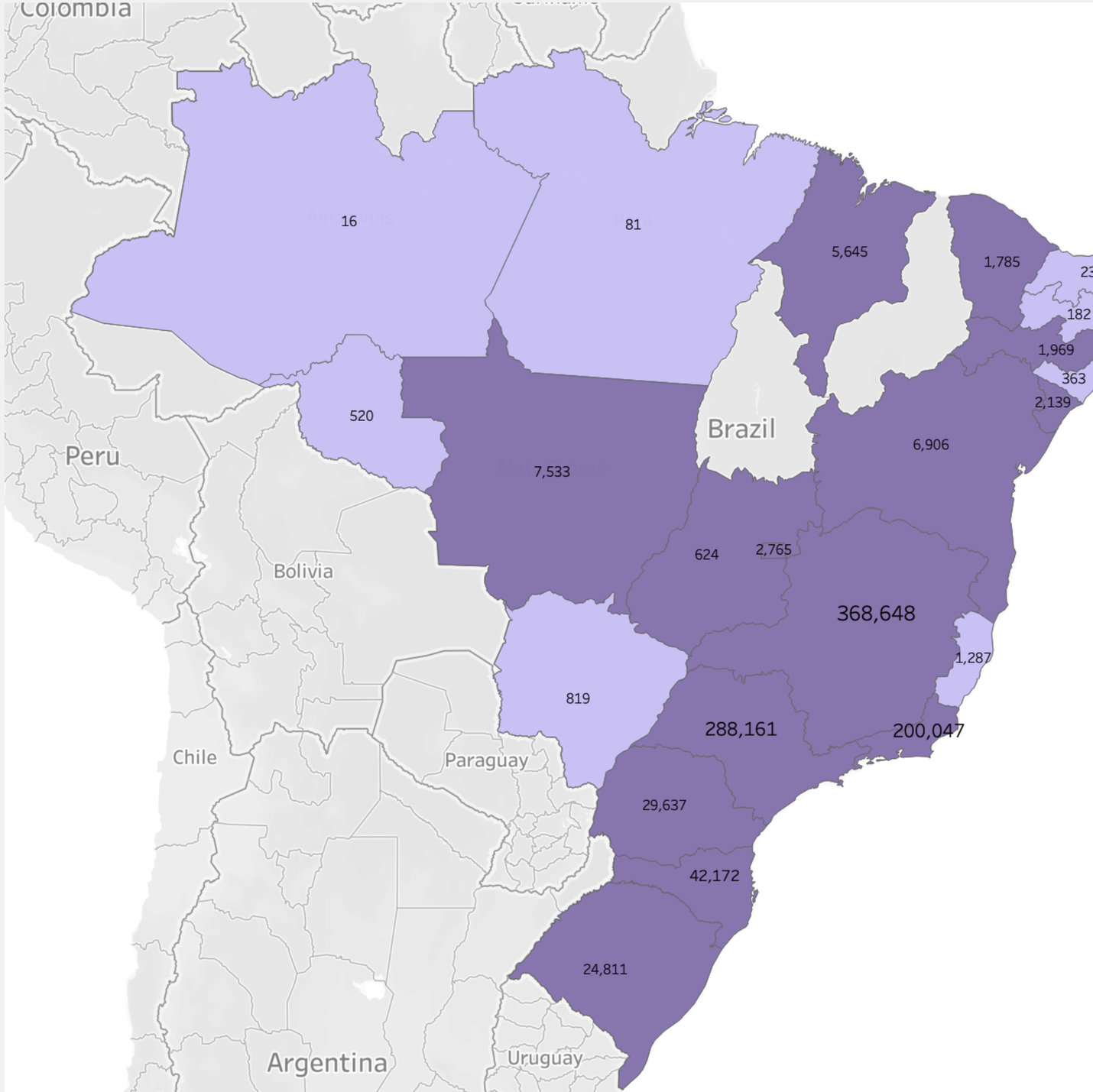
Determining the most popular product category and ascertaining which state had the highest sales revenue with the geospatial analysis.

What went wrong

Since the data lacked multiple numerical value columns I could not derive meaningful insight from the cluster analysis.

Lessons leaned

For a precise seasonal trend, the dataset should encompass a minimum of five years, enabling the capture of more nuanced and intricate seasonal fluctuations.



CONTACT ME

jennifersousa9@gmail.com

1-682-559-0073

jhttps://www.linkedin.com/in/jennifersousa5/