**Capstone Project**

**Table of Contents:**

* Introduction
* Ask phase
* Prepare phase
* Process phase
* Analyze phase
* Share phase
* Act Phase

**Objectives:**

* To analyze the best category
* To analyze the best sales on month
* To analyze the patterns in data
* To provide the meaningful visualizations

1. ***Introduction:***

About the Company:

It is a sports company in USA that provides different types of categories like Exercise and fitness, outdoor recreation, gymnastics, team sports, puzzles, etc. It provides these sports to different states in USA.

1. ***Ask Phase:***

Business task:

To show the patterns in transaction data. Making some decisions to know the which category is most purchased in USA.

1. **Pr*epare Phase:***

About the dataset:

I have downloaded dataset from GitHub. The data is about transactions. The dataset contains 5257 rows and 9 columns about transactions in USA.

The attributes in the dataset are:

**Txno**: transaction number

**Txndate**: date of transaction

**Cust no**: customer number

**Amount**: amount of transaction

**Category**: type of category

**Product**: product purchased

**City:** city of transaction

**State**: state of transaction

**Spend by**: type of spend of money

***Tools used:***

**Excel**:used for cleaning

**Python**: used for cleaning and analysis

**Tableau**: used for visualization

1. ***Process Phase:***

I have imported the required libraries in Python. I have imported the data from my

system to python. Then I checked there are any null values or not by using isnull function and through heat map. I checked for datatypes. I cleaned data by some functions in python.

Then I saved the cleaned data from python to csv.

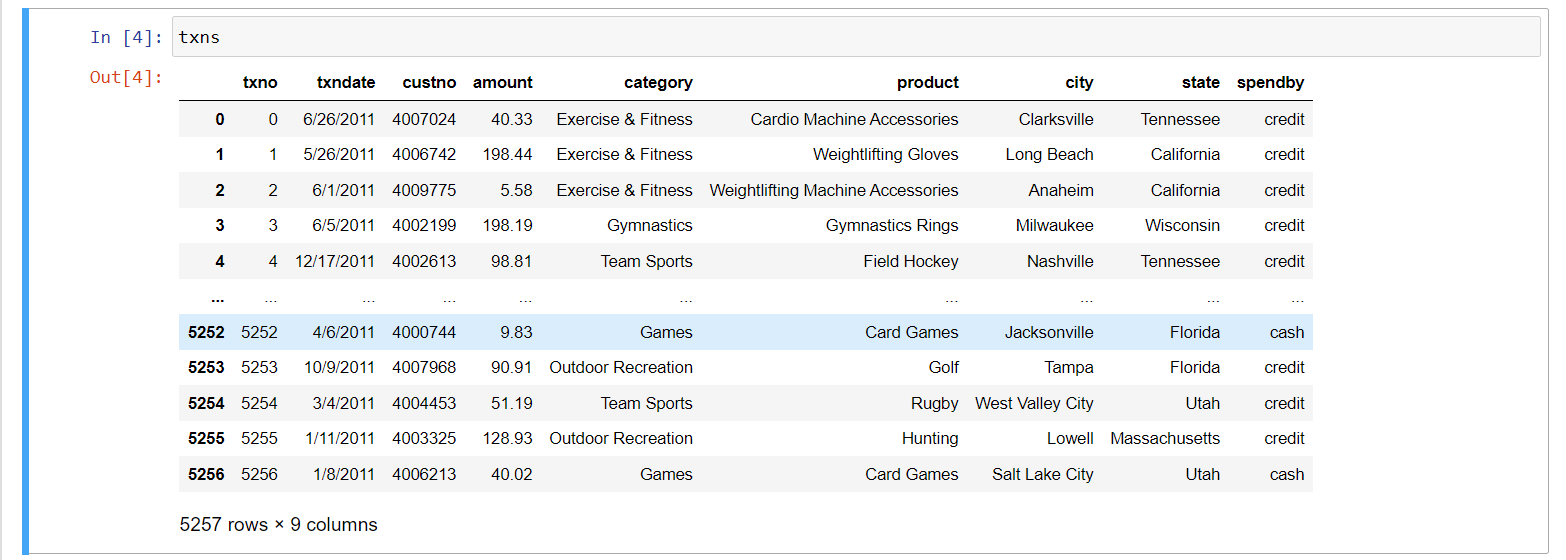
# importing the required libraries



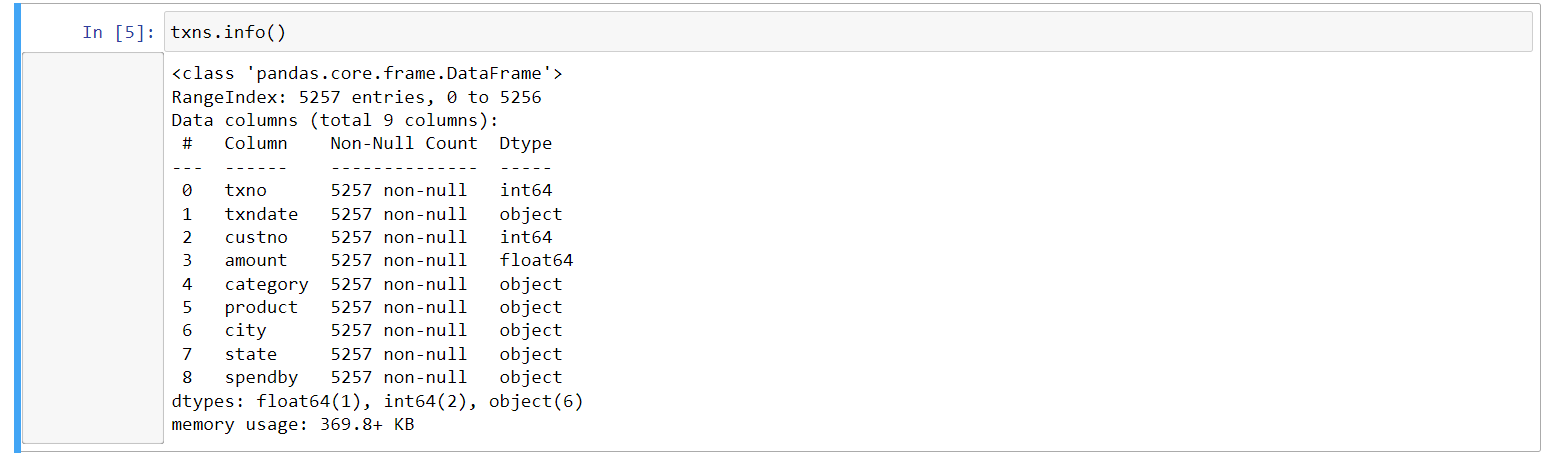
#importing the dataset to python



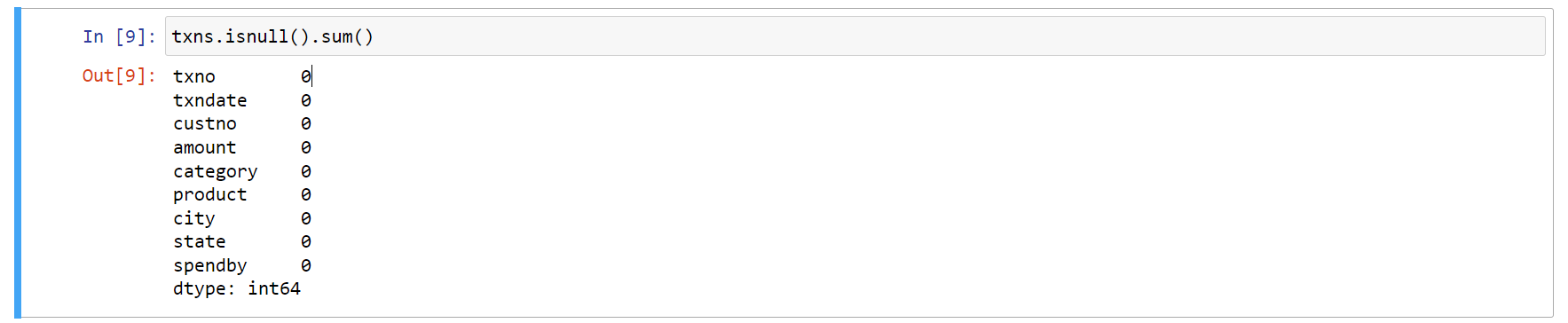
#Previewing the dataset



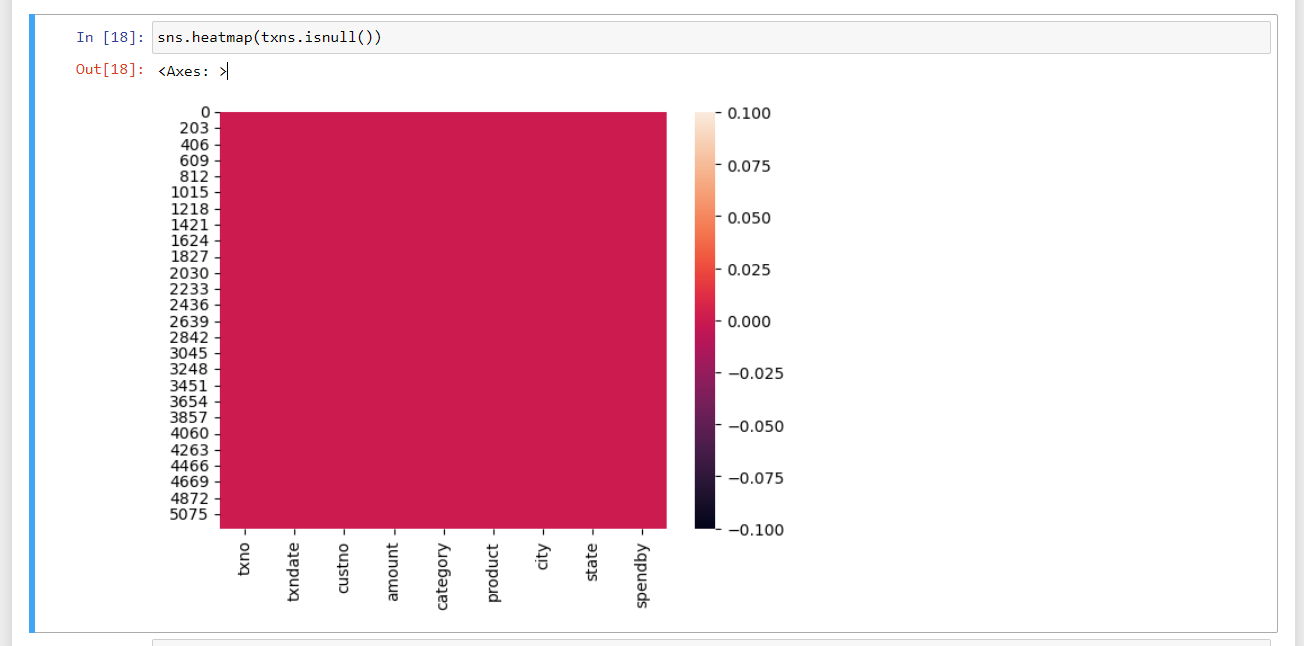
#checking info



#checking null values



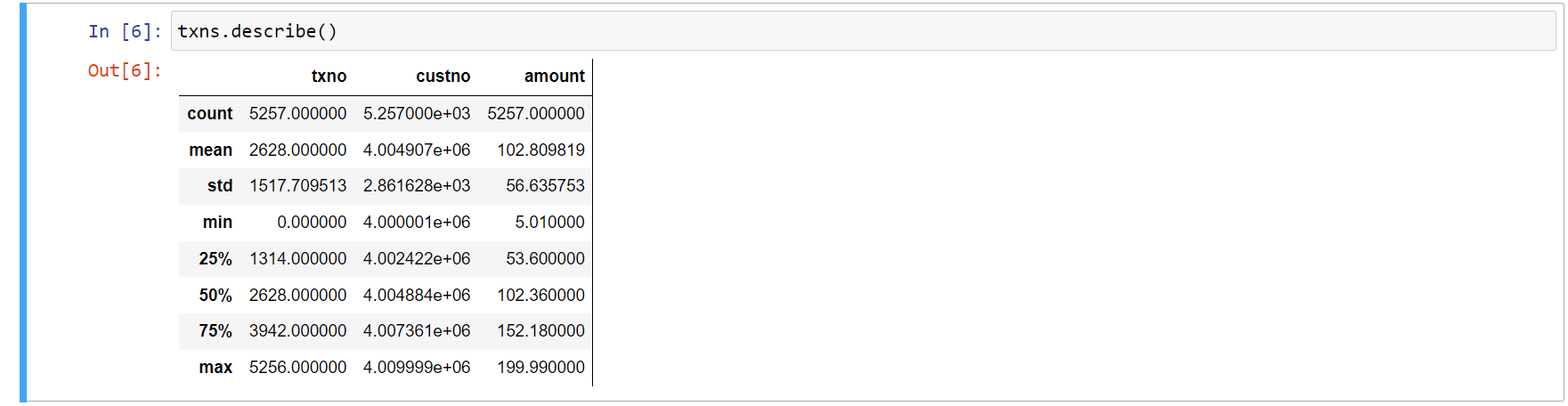
#checking null values through heat map



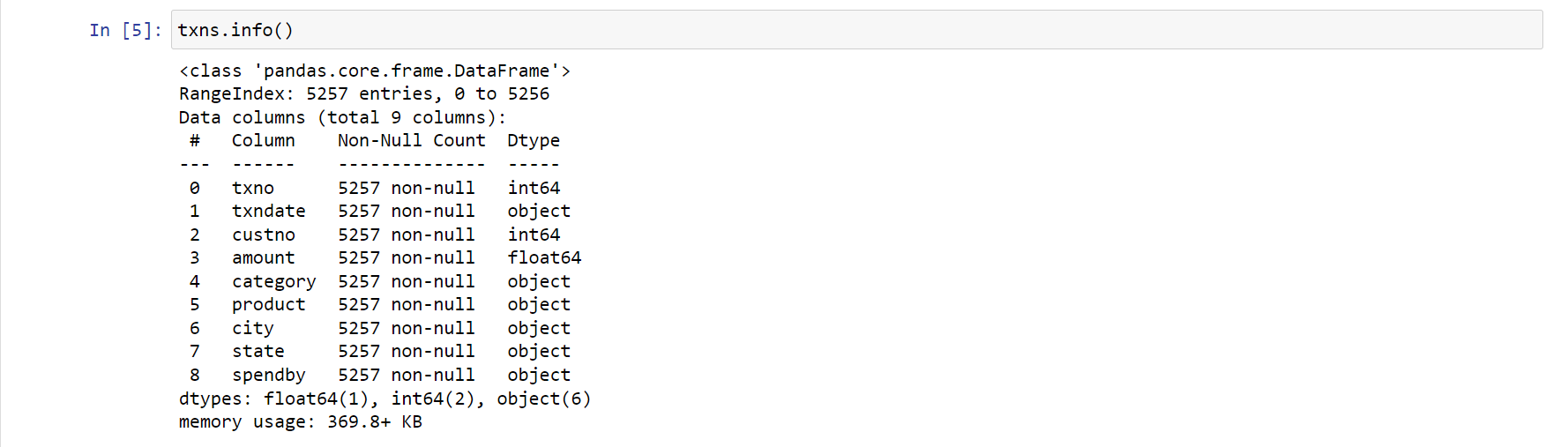
1. ***Analyze:***

In this phase I analyzed the data. I used sorting and filtering functions to find out the patterns. I also find out correlation between the variables. I did statistical analysis on attributes. I found how many unique elements are there in each attribute.

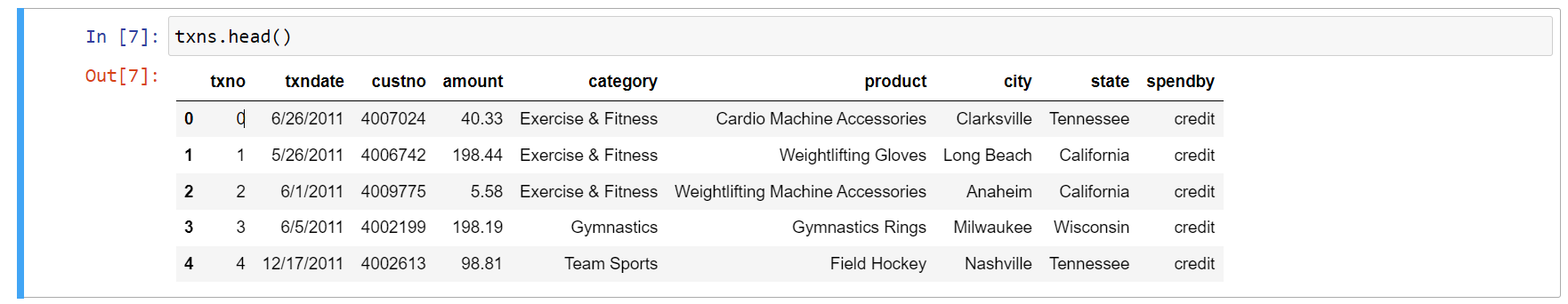
My Statistical Analysis is:



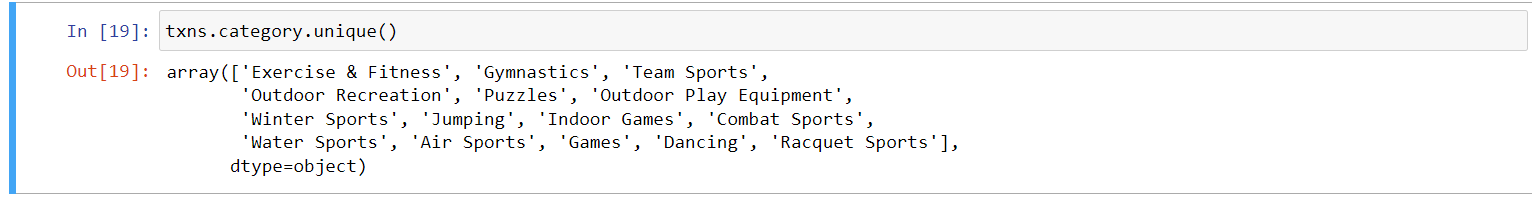
# info



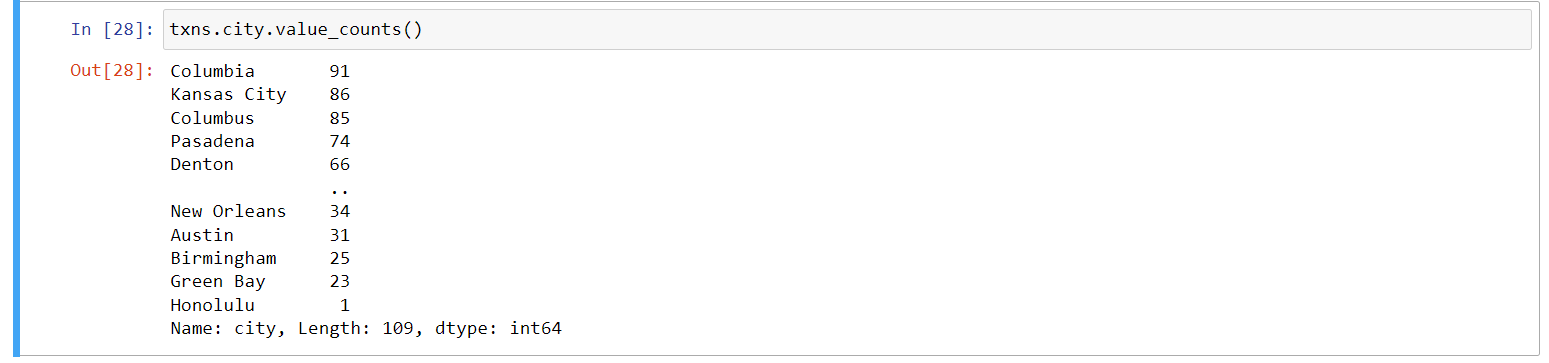
# using head function

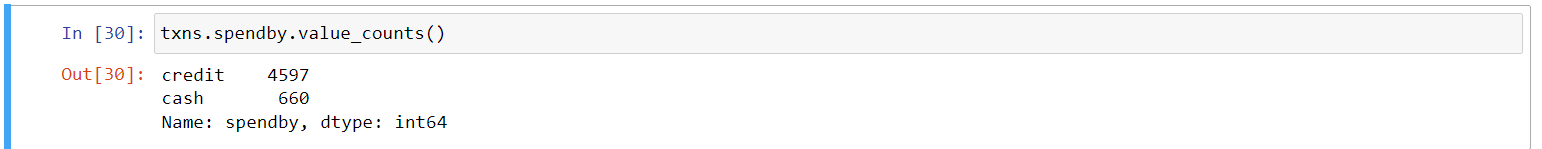


# Identifying unique

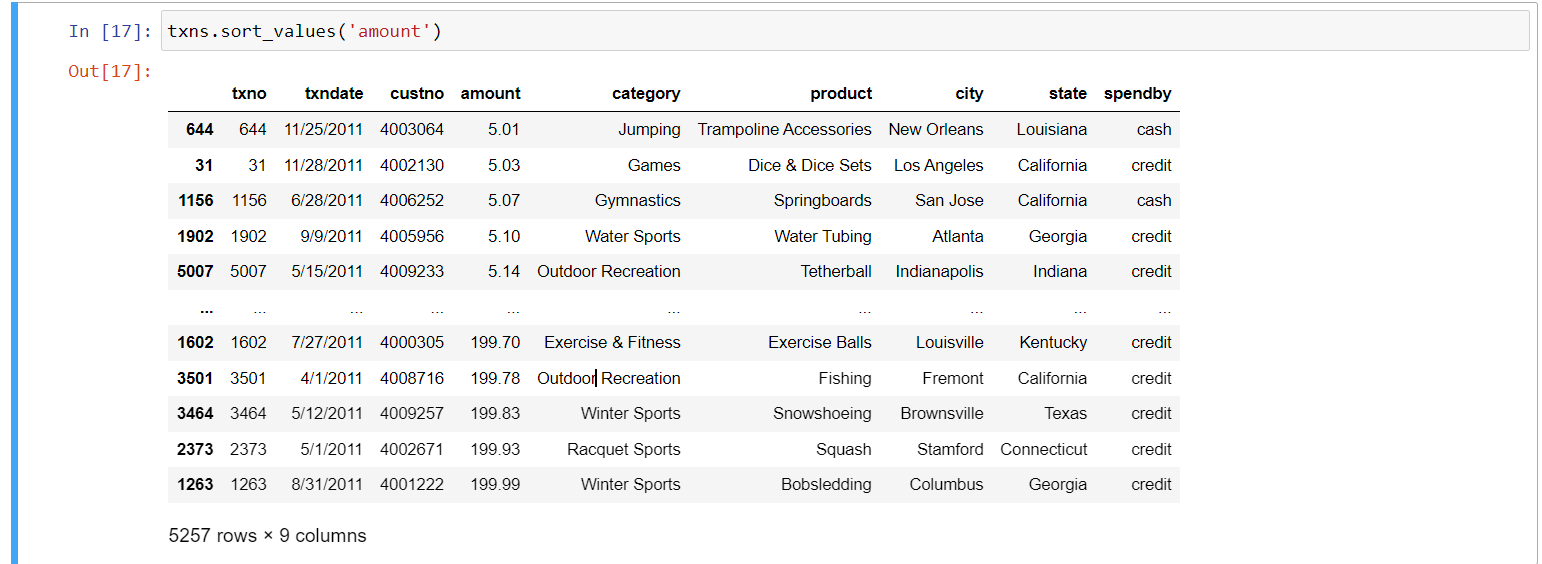


# counts

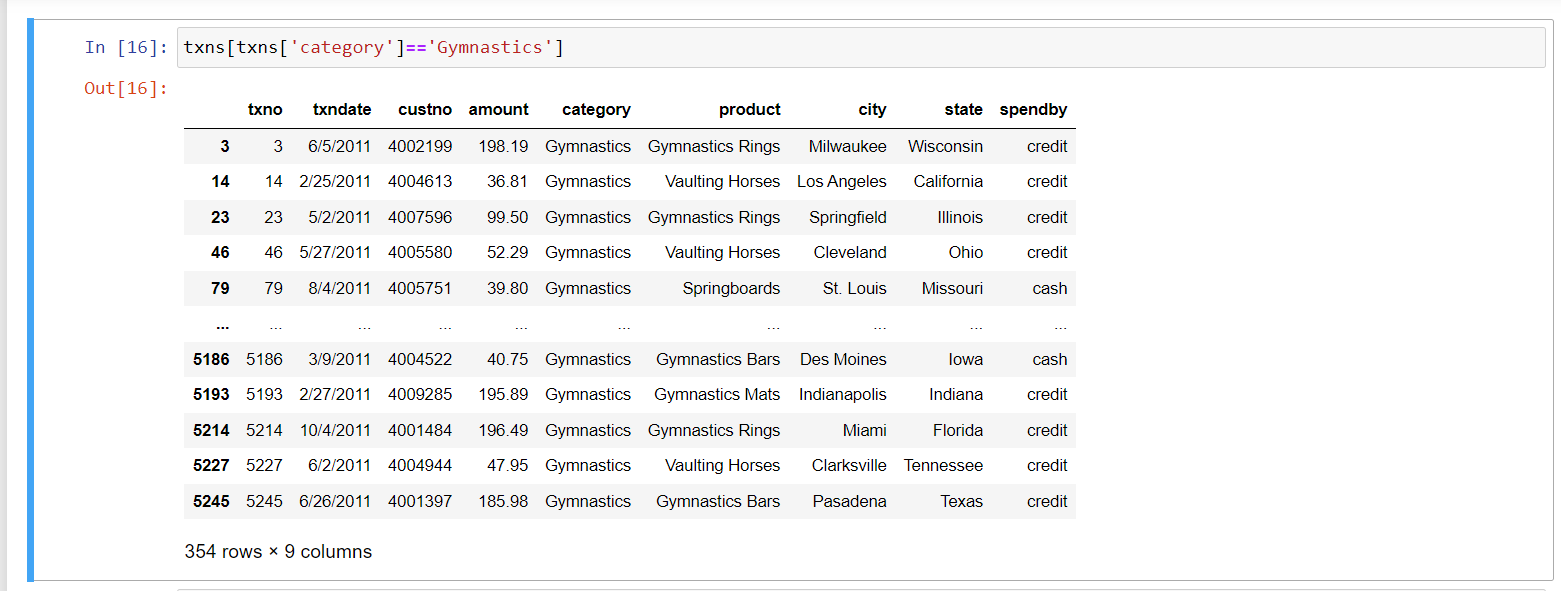




# sorting

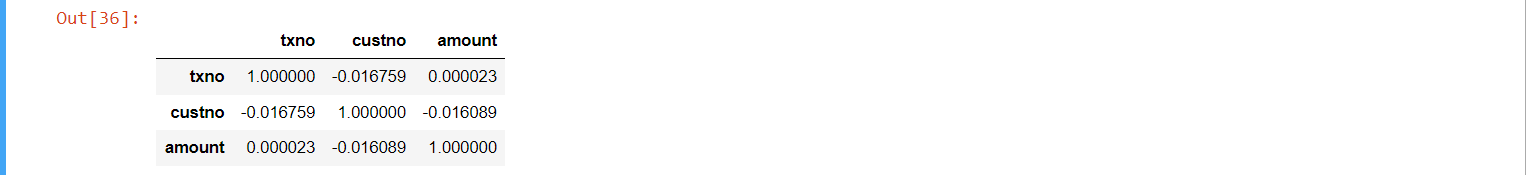


# Filtering

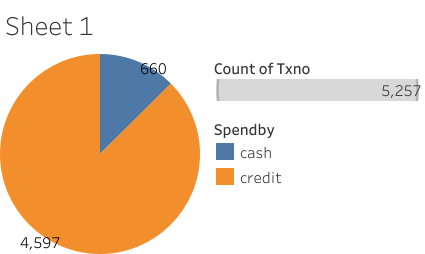


# finding Correlation



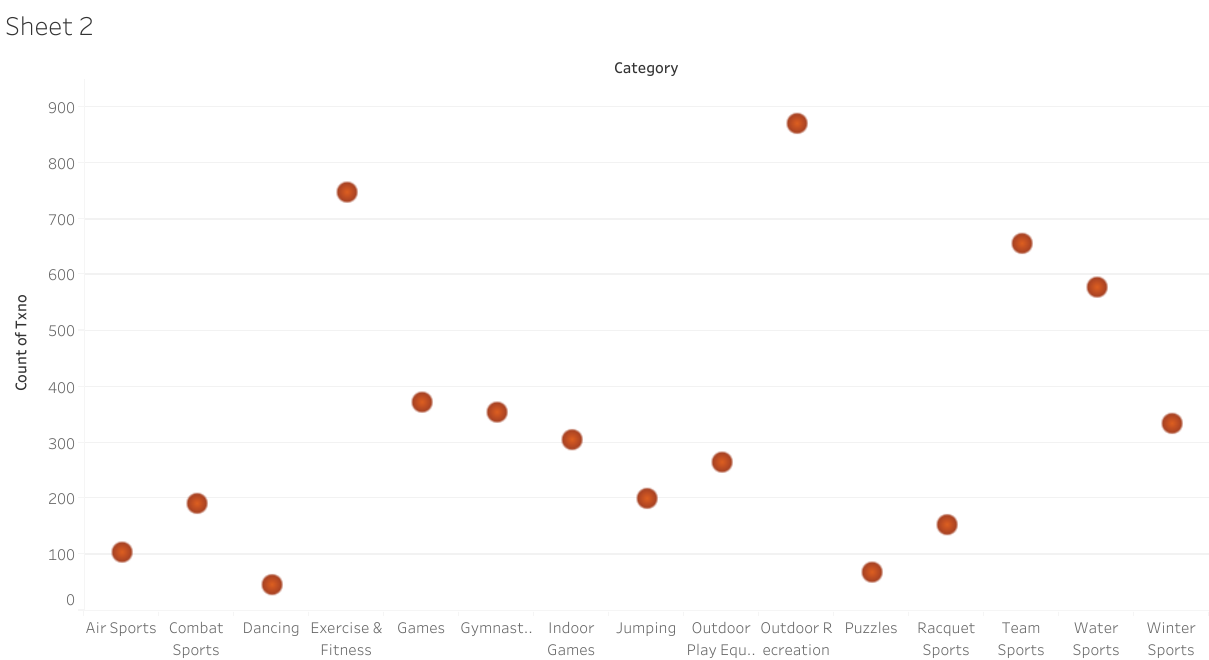


1. ***Share:***
2. finding how many cash and credit transactions



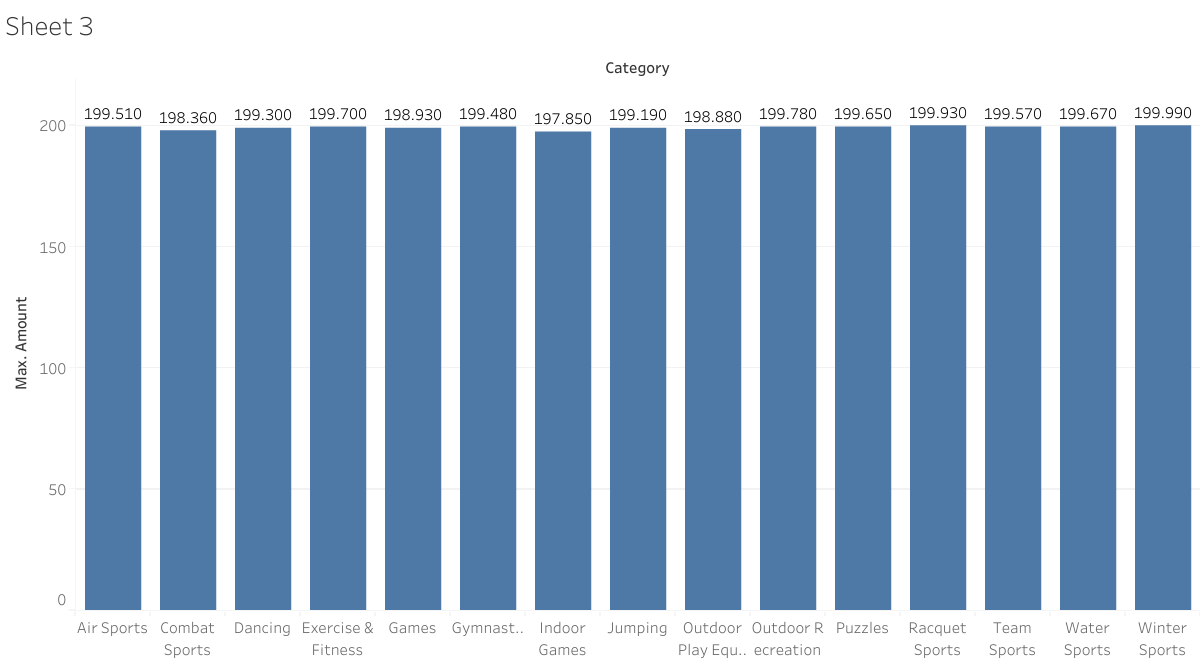
Most customers used credit transactions only. There are only 660 cash transactions and 4,597 credit transactions. Now-a-days most people are using credit transactions for secure payments.

1. highest category



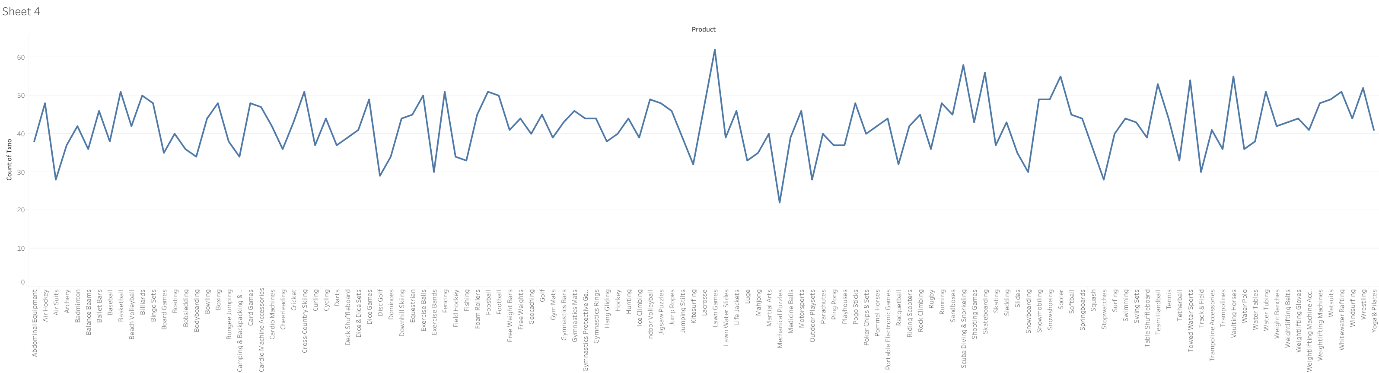
Most customers chosen outdoor recreation category. It refers to leisure activities that takes place in a natural setting and benefit the body. Examples of outdoor recreation are hiking, fishing and wild life viewing.

1. Maximum amount



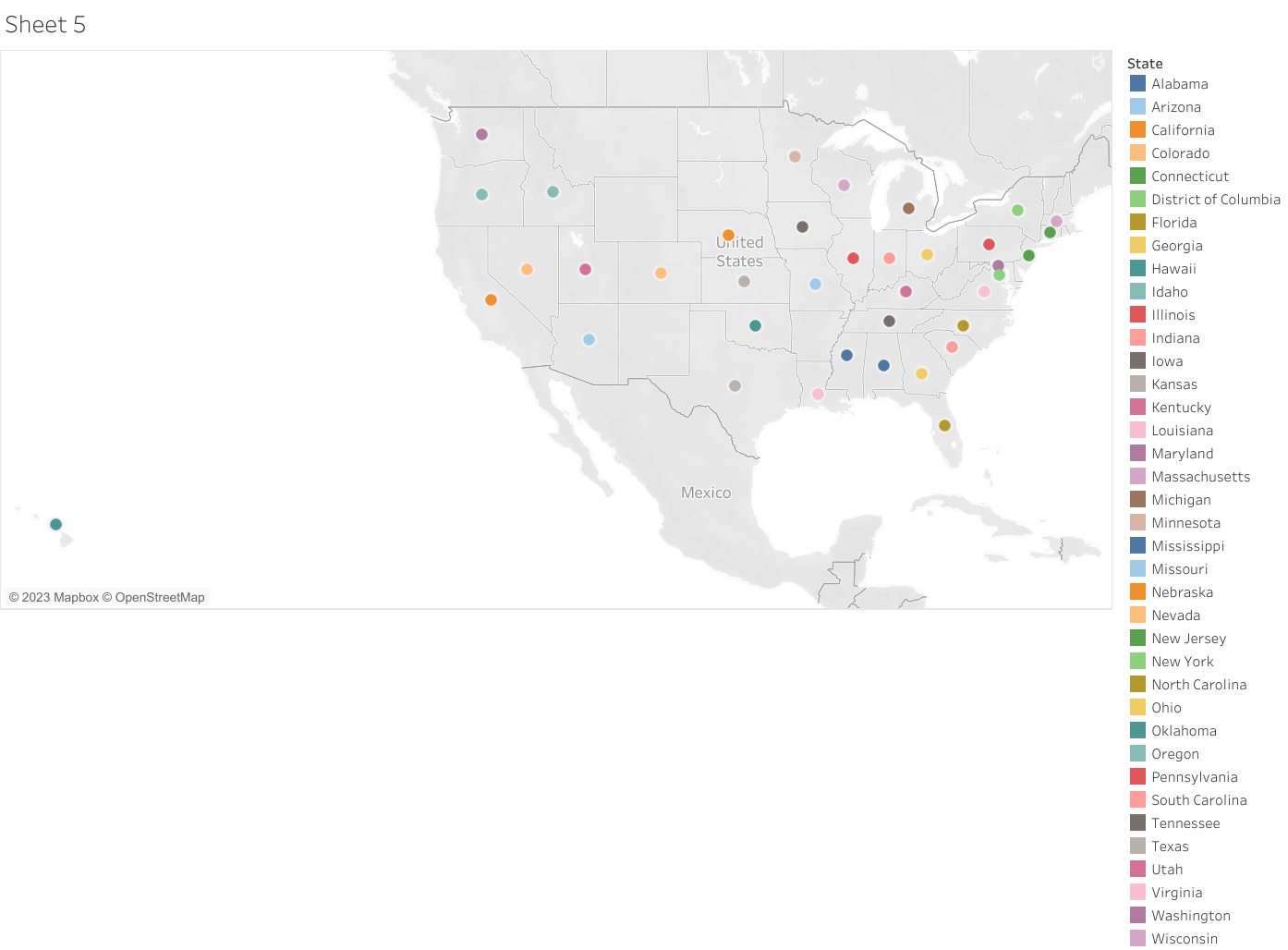
The winter sports have the highest amount. Because it is a seasonal sport. Winter sports include ice hockey, skiing, snowboarding, skating, etc. Such games were only played in cold areas during winter.

1. highest product



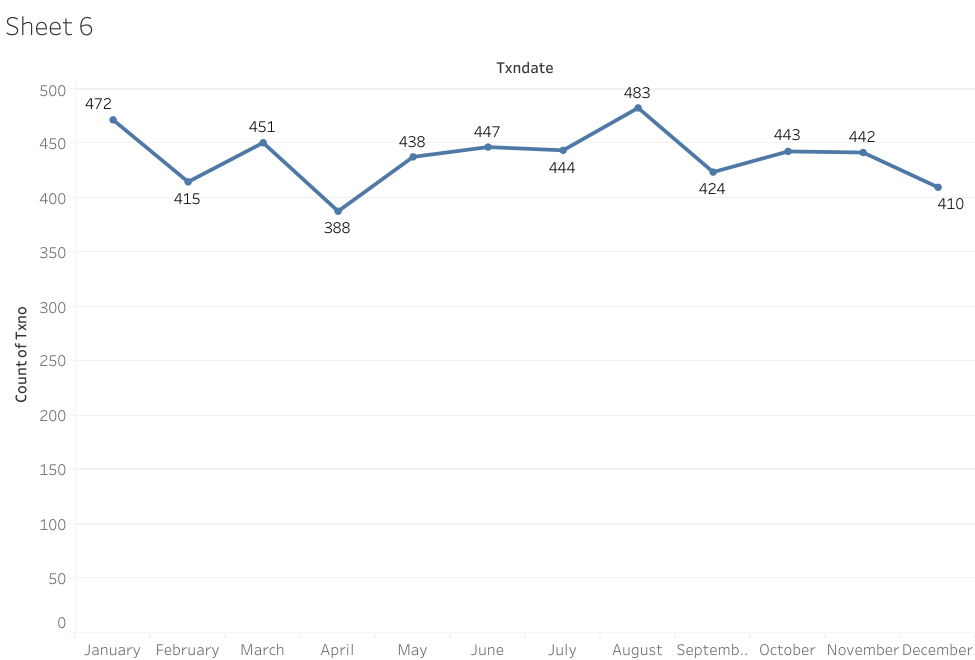
Most customers have chosen Lawn games under outdoor recreation category. Some of the most popular lawn games include croquet, tug-of-war and lawn bowls. These games are outdoor games played on lawn.

1. number of transactions in each state



By this map we can find out how many transactions in each state. California has the highest no of transactions that is 741 and Louisiana has the lowest no of transactions that is 34.

1. no of transactions for each month



By above line chart we can know the number of transaction for each month. August has the highest number of transactions and April has the lowest.

1. ***Act:***

**Conclusion:**

Here we performed exploratory analysis on transaction data of USA and looked into the most interesting factors that led to a sports company successfully running in the city. The code provided here can be easily understood and used to implement EDA on other similar datasets.