

# About Team

## Team Name

* Minions

## Team Member

Dokyeong Kwon

Seungwoo Park

Taeseung Lee

# About Project

## Project Summary

* Title
* Whiskey Recommendation System
* Goal
* Analyze the whiskey survey and recommend another Whiskey that the user would like by receiving one of the favorite whiskey.

## Project Plan(Briefly)

|  |  |
| --- | --- |
| Week | * About |
| Week 1 | Determine Project topics, Investigate Domains, Analyze data |
| Week 2 | Feature Selection, Data Cleaning and Preprocessing |
| Week 3 | Modeling and Testing |
| Week 4 | Web Application Development |

## Project Reference

* Python Syntax
* <https://docs.python.org/3/contents.html>
* Pandas API Reference
* <https://pandas.pydata.org/pandas-docs/stable/api.html>
* Visualizing the distribution of a dataset
  + <https://seaborn.pydata.org/tutorial/distributions.html>
  + <https://www.kaggle.com/residentmario/univariate-plotting-with-pandas>
* GitHub
* <https://github.com/orgs/minionsDoJo/dashboard>
* Google Drive
* <https://drive.google.com/drive/u/0/folders/1sotSK_wo2VQZNGV92Td9BA92ohz8ygCO>

# Data description

* 

## Index

* Index Numbering

## Name

* Name of whiskey bottle

## Category

* Whisky Category

## Review, Point

* Point marked by each reviewers

## Price

* Price of each bottle

## Currency

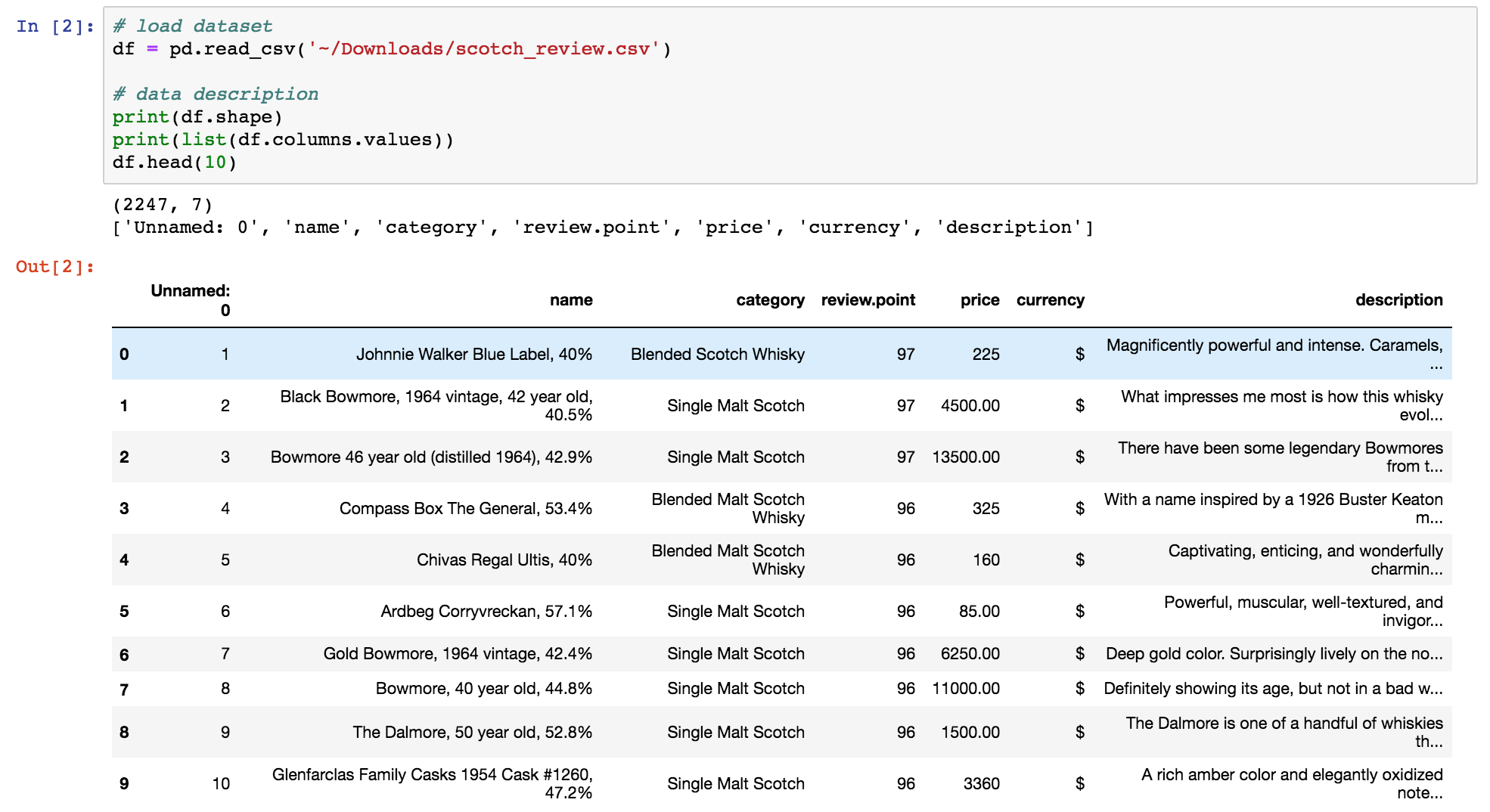
* Unit of price

## Description

* Descriptions of reviews
* Site : https://www.kaggle.com/koki25ando/22000-scotch-whisky-reviews

# Methods and Result

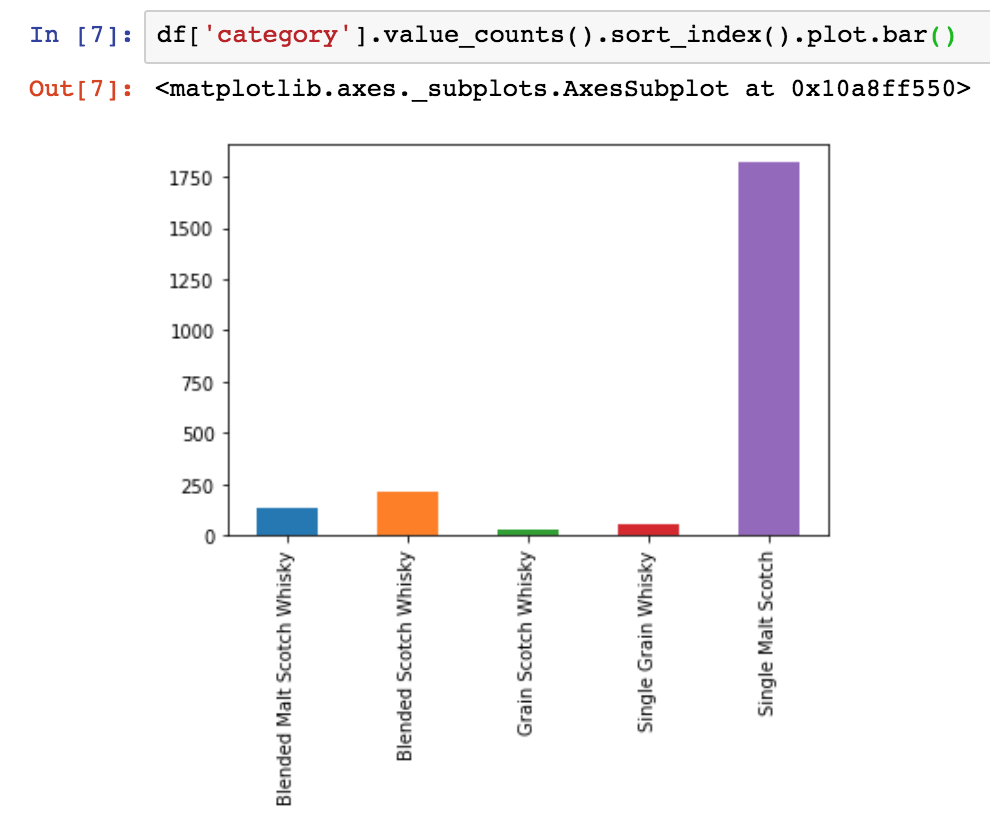
* 1. Methods
     1. Language : Python
     2. Libraries : Numpy, Pandas, Matplotlib
     3. Tools : Jupyter NoteBook. anaconda(package management)
  2. Result
     1. Shape, column names and top 10 values of Dataset ‘scotch\_review.csv’



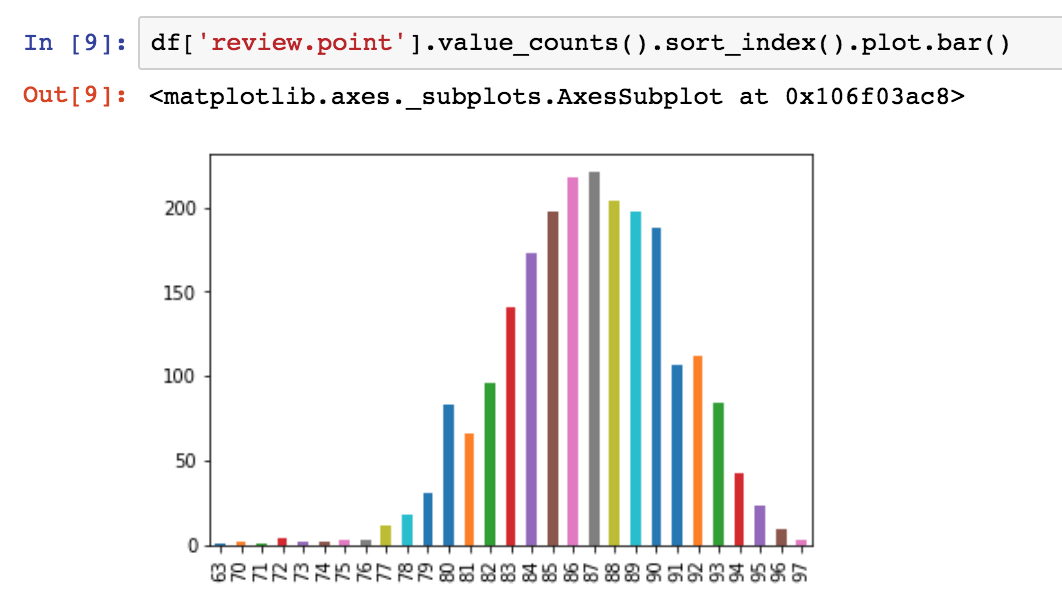
* + 1. Column name ‘Unnamed: 0’ and currency don’t need to used for our project.
       1. Values in Unnamed: 0 column just say their {index+1} and don’t used at any columns.
       2. There are only dollars in the currency column.



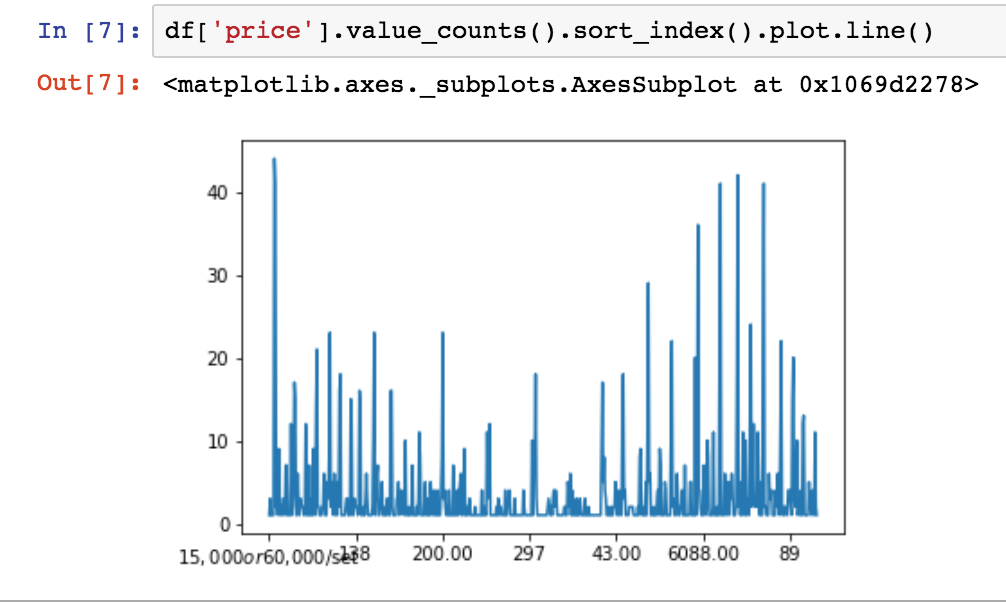
* + 1. The Category bar plot (x : categories, y : counts)



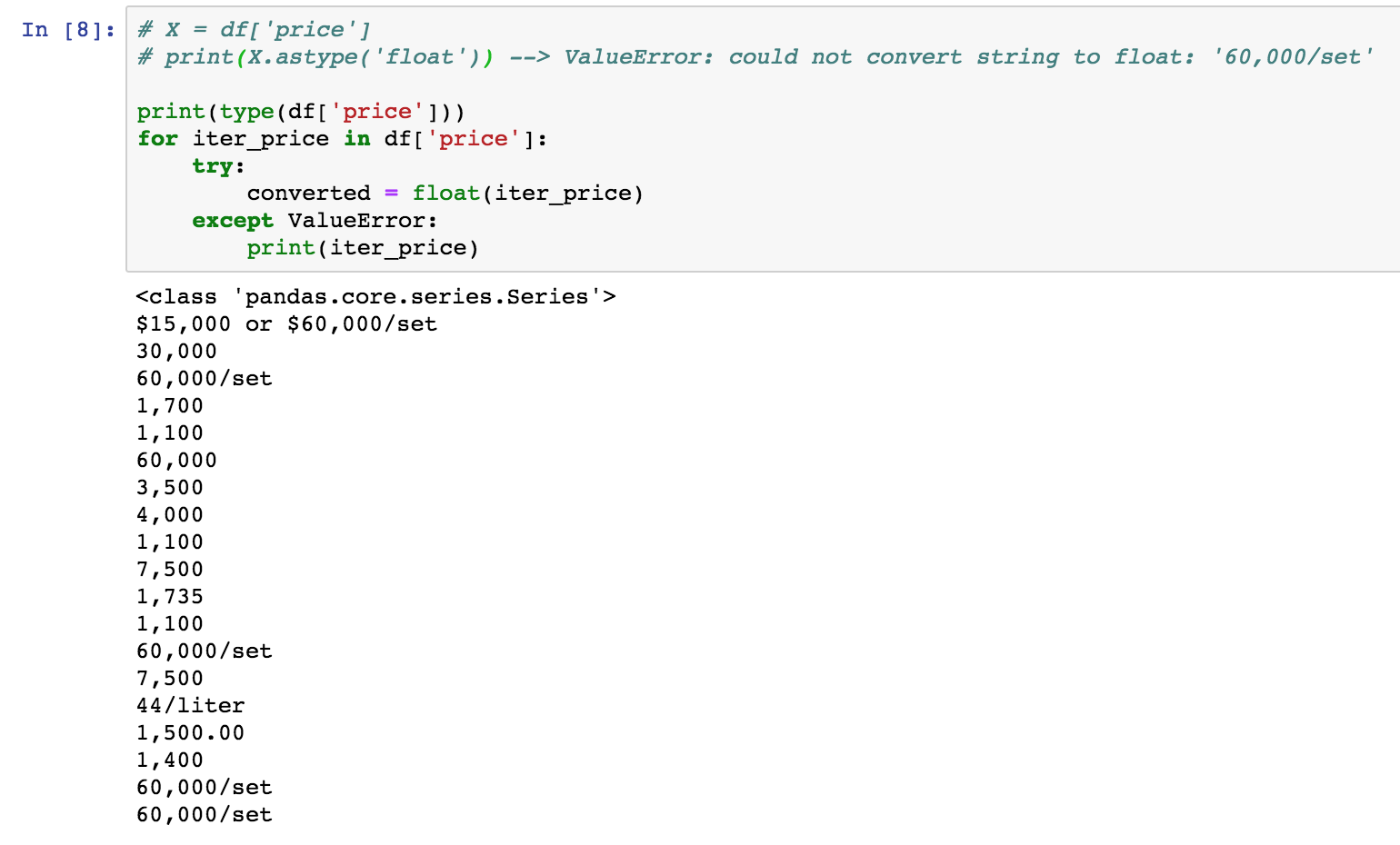
* + 1. The review point bar plot (x : review points, y : counts)



* + 1. The price line plot (x : prices, y : counts)



* 1. There are price data that cannot be converted to float or int64 (need to clean)



1. Conclusion

In this week, We decided our project topic and we analyzed about the data. And In future, We will cleaning data by removing useless columns and preprocessing about our data. Especially, due to we need the analysis about whiskey description, we will decide the method for text analysis.