Beer_data_analysis

November 26, 2025

[1]:
```
import pandas as pd
```

[4]:
```
df = pd.read_csv("BeerDataScienceProject.tar.bz2", compression='bz2', sep=',')
df.head()
```

[4]:

	beer_ABV	beer_beerId	beer_brewerId	beer_name	beer_style	review_appearance	review_palette	review_overall	review_taste	review_profileName	review_aroma	review_text	review_time
0	5.0	47986	10325	Sausa Weizen	Hefeweizen	2.5	2.0	1.5	1.5	stcules	1.5	A lot of foam. But a lot. In the smell some ba…	1234817823
1	6.2	48213	10325	Red Moon	English Strong Ale	3.0	2.5	3.0	3.0	stcules	3.0	Dark red color, light beige foam, average. In …	1235915097
2	6.5	48215	10325	Black Horse Black Beer	Foreign / Export Stout	3.0	2.5	3.0	3.0	stcules	3.0	Almost totally black. Beige foam, quite compac…	1235916604
3	5.0	47969	10325	Sausa Pils	German Pilsener	3.5	3.0	3.0	2.5	stcules	3.0	Golden yellow color. White, compact foam, quit…	1234725145
4	7.7	64883	1075	Cauldron DIPA	American Double / Imperial IPA	4.0	4.5	4.0	4.0	johnmichaelsen	4.5	According to the website, the style for the Ca…	1293735206

1. Rank the top 3 breweries which produce the strongest beers.

[82]:
```
breweries_producing_strongestbeer_df=df.groupby("beer_brewerId")["beer_ABV"].
↪mean().sort_values(ascending=False).head(3).reset_index()
breweries_producing_strongestbeer_df
```

[82]:
```
   beer_brewerId   beer_ABV
0           6513  19.228824
1            736  13.750000
2          24215  12.466667
```

2. Which year did beers enjoy the highest ratings?

[72]:
```
df['review_year']=pd.to_datetime(df['review_time']).dt.year
```

[90]:
```
year_with_highest_rating=df.groupby('review_year')["review_overall"].mean().
 ↪sort_values(ascending=False)
year_with_highest_rating
```

[90]:
```
review_year
1970    3.833197
Name: review_overall, dtype: float64
```

3. Based on the users' ratings, which factors are important among taste, aroma, appearance, and palette?

[104]:
```
important_factors_df=df.
 ↪groupby('beer_beerId')[["review_taste","review_aroma","review_appearance","review_palette",
 ↪mean()
important_factors_df.corr()
```

[104]:
```
                   review_taste  review_aroma  review_appearance  \
review_taste           1.000000      0.821956           0.659598
review_aroma           0.821956      1.000000           0.637400
review_appearance      0.659598      0.637400           1.000000
review_palette         0.736896      0.813106           0.647649
review_overall         0.809601      0.873737           0.614839

                   review_palette  review_overall
review_taste             0.736896       0.809601
review_aroma             0.813106       0.873737
review_appearance        0.647649       0.614839
review_palette           1.000000       0.747198
review_overall           0.747198       1.000000
```

From the above correlation between all review columns, "aroma", "palettee" and then "taste" is very important respectively for users as compared to nay other factor.

4. If you were to recommend 3 beers to your friends based on this data, which ones would you recommend?

[154]:
```
beers_df=df.
 ↪groupby("beer_name")[["review_aroma","review_appearance","review_palette"]].
 ↪mean().
 ↪sort_values(by=["review_aroma","review_appearance","review_palette"],ascending=False)
```

```
beers_df.head(3)
```

[154]:

beer_name	review_aroma	review_appearance	review_palette
Blueberry Hefeweizen	5.0	5.0	5.0
Date Night With Jumbo Love	5.0	5.0	5.0
Dry Hopped Abominable Ale	5.0	5.0	5.0

5. Which beer style seems to be the favourite based on the reviews written by users? How does written reviews compare to overall review score for the beer style?

[136]:
```
favourite_beer_df=df.groupby(by="beer_style")["review_overall"].mean().
↪sort_values(ascending=False).head(1).reset_index()
favourite_beer_df
```

[136]:

	beer_style	review_overall
0	Gueuze	4.140952