# SUMMARY

## Q1: Which brewery produces the strongest beers by abv ?

Answer:

In [189…

```
beer_reviews_by_brewery_stats = beer_reviews_copy[["brewery_name", "beer_abv"]].groupby
beer_reviews_by_brewery_stats.columns = beer_reviews_by_brewery_stats.columns.droplevel
beer_reviews_by_brewery_stats.sort_values(by=["max", "var", "mean"], ascending=[False,
```

Out[189…

| brewery_name | min | mean | max | var | count |
|---|---|---|---|---|---|
| **Schorschbräu** | 4.900 | 19.229 | 57.700 | 151.962 | 34 |
| **BrewDog** | 0.500 | 9.905 | 41.000 | 39.208 | 4033 |
| **De Struise Brouwers** | 2.000 | 10.763 | 39.000 | 10.425 | 3866 |
| **Hair of the Dog Brewing Company / Brewery and Tasting Room** | 3.200 | 9.407 | 29.000 | 5.374 | 3769 |
| **Boston Beer Company (Samuel Adams)** | 4.000 | 6.415 | 27.000 | 8.895 | 38812 |
| **…** | … | … | … | … | … |
| **Yellow Rose Brewing Company** | 0.000 | 0.000 | 0.000 | 0.000 | 0 |
| **Yozgut Bira Fabrikasi / Tekel Birasi** | 0.000 | 0.000 | 0.000 | 0.000 | 0 |
| **Yuiga Doxon Ji Beer** | 0.000 | 0.000 | 0.000 | 0.000 | 0 |
| **Zea Rotisserie & Brewery** | 0.000 | 0.000 | 0.000 | 0.000 | 0 |
| **Zerodegrees Bristol** | 0.000 | 0.000 | 0.000 | 0.000 | 0 |

5742 rows × 5 columns

http://cms.worlds-strongest-beer.com/wp-content/uploads/2014/01/Schorschbock-57-033-quadrat.jpg

### How I got there:

- Data aggregation
- Sorting by Max Desc, Var Asc, Mean Desc

## Q2: If you had to pick 3 beers to recommend to someone, how would you approach the problem ?

Examples:

In [112…

```
beer_content_based_recommender(("Vecchio Birraio", "Hefeweizen", "Sausa Weizen")).sort_
```

```
initial beer index: 0
len scores: 3
recommended beers indexes: [1, 2, 3]
```

Out[112…

| | brewery_name | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|---|
| **0** | Vecchio Birraio | Foreign / Export Stout | Black Horse Black Beer | 3.000 | 1 | 0.027 |

| | brewery_name | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|---|
| **1** | Vecchio Birraio | English Strong Ale | Red Moon | 3.000 | 1 | 0.027 |
| **2** | Vecchio Birraio | German Pilsener | Sausa Pils | 3.000 | 1 | 0.027 |

In [113…
```
beer_content_based_recommender(("The Defiant Brewing Company", "Pumpkin Ale", "The Hors
```

```
initial beer index: 65753
len scores: 3
recommended beers indexes: [65709, 57392, 14767]
```

Out[113…

| | brewery_name | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|---|
| **0** | Voodoo Brewing Company | Pumpkin Ale | Headless Horseman | 3.731 | 13 | 0.262 |
| **1** | The Defiant Brewing Company | Pumpkin Ale | Headless Horseman Pumpkin Ale | 3.571 | 7 | 0.161 |
| **2** | Mishawaka Brewing Company | American Amber / Red Ale | The Fifth Horseman | 4.000 | 1 | 0.027 |

In [114…
```
beer_content_based_recommender(("Warsteiner Brauerei", "German Pilsener", "Warsteiner P
```

```
initial beer index: 22531
len scores: 3
recommended beers indexes: [22577, 22587, 22578]
```

Out[114…

| | brewery_name | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|---|
| **1** | Warsteiner Brauerei | Märzen / Oktoberfest | Warsteiner Premium Oktoberfest | 3.552 | 165 | 0.819 |
| **2** | Warsteiner Brauerei | Fruit / Vegetable Beer | Warsteiner Premium Orange | 2.500 | 6 | 0.141 |
| **0** | Warsteiner Brauerei | Light Lager | Warsteiner Premium Light | 3.000 | 4 | 0.099 |

**How I got there:**

- Concatenation of text fields
- TFIDF -- Term Frequency * Inverse Document
  - Term Frequency ==
    – Number of occurrences of a term in the document (can be a simple count)
  - Inverse Document Frequency ==
    – How few documents contain this term or:
    – log (#documents / #documents with term)
- Computing Cosine similiarity:

  cosine = linear_kernel(tfidf_matrix, tfidf_matrix, dense_output=True)
- Filtering cosine dataset by indices of input dataset to get a list of most similiar beers

## Q3: What are the factors that impacts the quality of beer the most ?

**Answer:**

I hoped to get Factors that impacts the quality of beer the most from Feature Importance metrics. Unfortunately getting Feature Importance when you develop a model with a use of sklearn piplines is a bit tricky

**How I got there:**

- Imputation:
  - numeric:
    - SimpleImputer(strategy="median")
  - character
    - Most frequent
- Feature Engeneering:
  - Deviations defined as:
    - deviation_feature = ((dataframe[feature] - category_mean) / category_std).rename(category_feature)
  - Dummy Variables for:
    - "beer_style"
- Variables selection methods used:
  - VarianceThreshold:
    - threshold=(.6 * (1 - .6))) == 0.24
    - Variables of variance lower than 0.24 were removed from process of fitting classifier
  - SelectKBest(f_classif, k = 3)
    - "Compute the ANOVA F-value for the provided sample."
- Scaling of imputes data:
  - StandardScaler()
    - Devietions derived during Feature Enginnering are of diffrent scale than review Ratings
- Classifier RandomForestClassifier:

In [188…

```
print(score_df)
```

```
   accuracy  precision  recall    f1   auc
0     0.785      0.775   0.818 0.796 0.784
```

**...however no Feature Importance yet**

## Q4: I enjoy a beer which aroma and appearance matches the beer style. What beer should I buy ?

Answer:

In [267…

```
reviews_by_beer_style_mean_count = beer_reviews_copy[["review_overall", "beer_style", "
reviews_by_beer_style_mean_count.columns = reviews_by_beer_style_mean_count.columns.dro
```

```
reviews_by_beer_style_mean_count = reviews_by_beer_style_mean_count.reset_index()
reviews_by_beer_style_mean_count
```

Out[267...

| | beer_style | beer_name | mean | count |
|---|---|---|---|---|
| **0** | Altbier | "Alt"ered State | 3.500 | 2 |
| **1** | Altbier | "Nein Toll Bier" German Alt | 4.000 | 1 |
| **2** | Altbier | 'Sconnie Rustic Trail Amber | 2.250 | 2 |
| **3** | Altbier | (512) Alt | 4.091 | 11 |
| **4** | Altbier | 2010 Cleveland Beer Week Collaboration Altbier... | 3.000 | 5 |
| **...** | ... | ... | ... | ... |
| **58887** | Witbier | ZomerWit | 2.500 | 1 |
| **58888** | Witbier | Zomerbier | 4.000 | 11 |
| **58889** | Witbier | Zomewhit | 4.500 | 1 |
| **58890** | Witbier | Zon Jager Wit | 4.000 | 1 |
| **58891** | Witbier | ZÔN | 3.829 | 208 |

58892 rows × 4 columns

In [262...

```
q1 = """
SELECT \
beer_style \
,beer_name \
,mean \
,count \
,((count / (count + (SELECT 2000 AS MinCount)) * mean) + (count / (count + (SELECT 2000
FROM reviews_by_beer_style_mean_count \
ORDER BY Score desc
"""
scores = ps.sqldf(q1, locals())
scores
```

Out[262...

| | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|
| **0** | American Double / Imperial IPA | 90 Minute IPA | 4.146 | 3290 | 0.310 |
| **1** | Russian Imperial Stout | Old Rasputin Russian Imperial Stout | 4.174 | 3111 | 0.299 |
| **2** | American IPA | Sierra Nevada Celebration Ale | 4.169 | 3000 | 0.291 |
| **3** | American IPA | Two Hearted Ale | 4.330 | 2728 | 0.272 |
| **4** | American Double / Imperial IPA | Stone Ruination IPA | 4.162 | 2704 | 0.270 |
| **...** | ... | ... | ... | ... | ... |
| **58887** | Witbier | Zeeuwsche Witte | 3.500 | 1 | 0.000 |
| **58888** | Witbier | Zoe's Belgian Wit | 4.000 | 1 | 0.000 |
| **58889** | Witbier | ZomerWit | 2.500 | 1 | 0.000 |
| **58890** | Witbier | Zomewhit | 4.500 | 1 | 0.000 |

| | beer_style | beer_name | mean | count | Score |
|---|---|---|---|---|---|
| **58891** | Witbier | Zon Jager Wit | 4.000 | 1 | 0.000 |

58892 rows × 5 columns

In [265…

```
scores['rank'] = scores.groupby(["beer_style"])['Score'].rank(method='first')
scores.sort_values(by=["rank"])
```

Out[265…

| | beer_style | beer_name | mean | count | Score | rank |
|---|---|---|---|---|---|---|
| **56133** | Oatmeal Stout | 'Fèileadh Air Teine' - Scottish Oatmeal Stout | 3.500 | 1 | 0.000 | 1.000 |
| **54885** | Lambic - Unblended | Amarillo Lambik | 4.000 | 1 | 0.000 | 1.000 |
| **52039** | Flanders Oud Bruin | Barrel-Aged Belgian Brown | 4.000 | 1 | 0.000 | 1.000 |
| **57275** | Schwarzbier | Černá Hora Granat | 3.500 | 1 | 0.000 | 1.000 |
| **40853** | American Brown Ale | 15.0 | 4.000 | 1 | 0.000 | 1.000 |
| **...** | ... | ... | ... | ... | ... | ... |
| **19** | American IPA | HopDevil Ale | 4.115 | 2302 | 0.240 | 3163.000 |
| **12** | American IPA | 60 Minute IPA | 4.173 | 2475 | 0.253 | 3164.000 |
| **7** | American IPA | Stone IPA (India Pale Ale) | 4.258 | 2575 | 0.261 | 3165.000 |
| **3** | American IPA | Two Hearted Ale | 4.330 | 2728 | 0.272 | 3166.000 |
| **2** | American IPA | Sierra Nevada Celebration Ale | 4.169 | 3000 | 0.291 | 3167.000 |

58892 rows × 6 columns

How I got there:

- Custom Score base on "review_overall"
- Ranking ["beer_style"] By Custom Score
- Sorting of Custom Score to get best beers among all ["beer_styles"]
- Filtering ["beer_styles"] by value of interest