Whisky pricing: A dram good case study
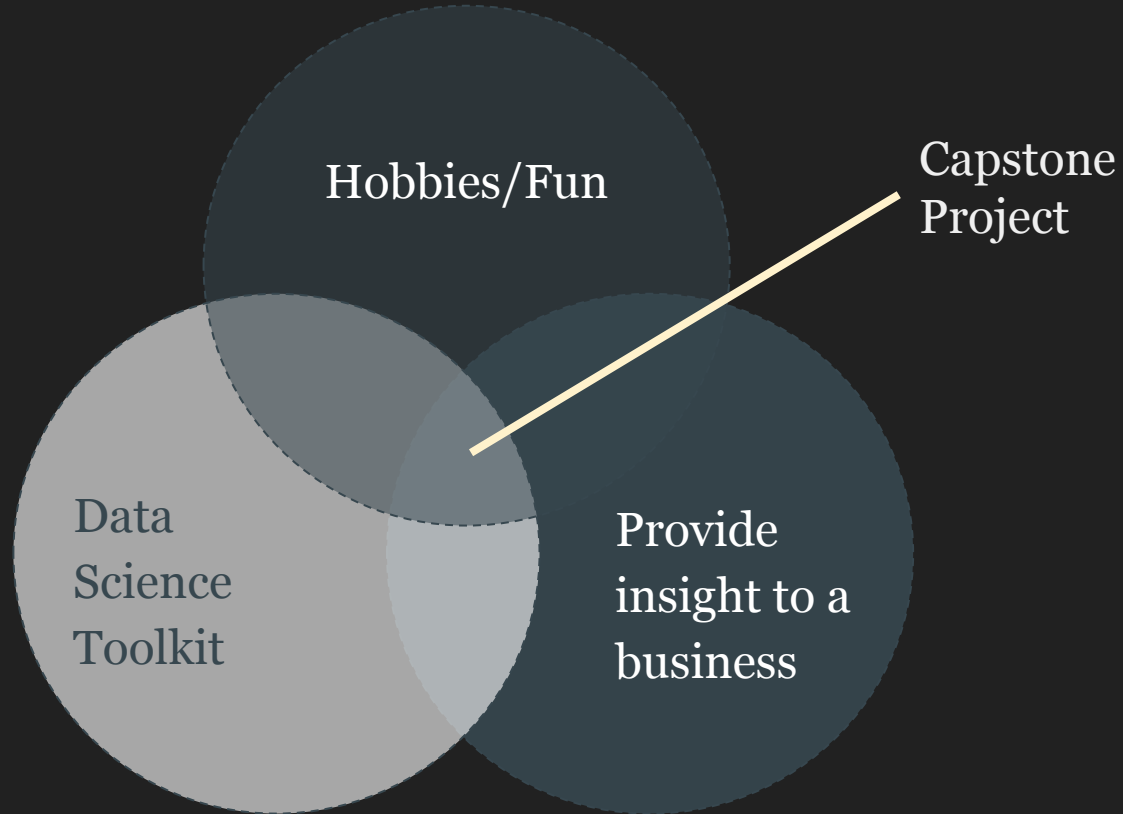
Anirudh Kashyap

General Assembly | 12/22/2017 | Capstone Project | The Whisky Exchange

Motivation

Hobbies/Fun

Capstone Project

Data Science Toolkit

Provide insight to a business

What factors affect the price of a whisky?

Contents

The Whisky Exchange (TWE)

- Spirits Retailer of the Year

- Worldwide Delivery to 55 countries

- Based out of London

- Value & Rare Malts

- First & fastest to grant permission

- US Liquor laws are confusing

Case Scenario

Data Collection

HIGHLAND PARK THE DARK 17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS	FACTS

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

Distillery

HIGHLAND PARK THE DARK
17 YEAR OLD

70cl / 52.9%

Island Single Malt Scotch Whisky
Distillery Bottling

DETAILS	FACTS

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

Whisky Name & Type

HIGHLAND PARK THE DARK 17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS | **FACTS**

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

HIGHLAND PARK

↑⌁⌁⊁ – THE DARK
SINGLE MALT SCOTCH WHISKY

17 YEAR OLD
1 of 28,000 bottles
PRODUCT OF SCOTLAND

HIGHLAND PARK
THE DARK 17 YEAR OLD
single malt scotch whisky alc.52.9%vol.700ml

ABV &
Volume

HIGHLAND PARK THE DARK
17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS	FACTS

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

Age

HIGHLAND PARK THE DARK 17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS	FACTS

Bottler
Distillery Bottling

Age
17 Year Old

No of Bottles
28000

Country
Scotland

Region
Island

Edition, Location

Other information

- Vintage (Year of release)

- Whisky type (Single Malt/Blended Malt etc.)

- Cask, Color, # of Reviews

- Price

Challenges

MONKEY SHOULDER

70cl / 40%

Blended Malt Scotch Whisky

DETAILS | FACTS

Country
Scotland

Colouring
Yes

HIGHLAND PARK THE DARK 17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS | FACTS

Bottler
Distillery Bottling

Age
17 Year Old

No of Bottles
28000

Country
Scotland

Region
Island

PORT ELLEN 1979
37 Year Old 17th Release Special Releases 2017

70cl / 51%

*Islay Single Malt Scotch Whisky
Distillery Bottling*

DETAILS | FACTS

Bottler
Distillery Bottling

Series
Annual Release

Age
37 Year Old

Vintage
1979

Bottling Date
2017

No of Bottles
2988

Country
Scotland

Region
Islay

Missing Values

Handling NaN (null) values

- Deductive Imputation

  - External Data

Distillery	Location	Region	Year closed
Pittyvaich	Dufftown	Speyside	1993
Port Charlotte	Port Charlotte	Islay	1929
Port Ellen	Port Ellen	Islay	1983
Rosebank	Falkirk	Lowlands	1993

https://en.wikipedia.org/wiki/List_of_whisky_distilleries_in_Scotland

Handling NaN values

- Deductive Imputation

  - Natural Language Processing (NLP)

HIGHLAND PARK THE DARK 17 YEAR OLD

70cl / 52.9%

*Island Single Malt Scotch Whisky
Distillery Bottling*

DETAILS	FACTS

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

Description (TWE / Bottle)

Whisky Description (NLP)

The first in a diptych that celebrates the seasons on the Isle of Orkney, where Highland Park is made. This bottle, The Dark, focuses on the autumn and winter seasons, while The Light – due to be released in 2018 – will symbolise spring and summer.

The Dark is a 17-year-old single malt that has been matured in sherry casks, giving it aromas of dried fruits, nuts and herbs that continue into the palate, where they are joined by distinctive notes of smoky peat.

The Dark has been bottled in a limited edition of 28,000.

Handling NaN values

- Deductive Imputation

    - Fill NaN (Back-fill-Front-fill)

    - Dropping columns with >90% NaN

Model Fitting

The distribution of whisky prices on TWE

Whisky prices range from under $50 for everyday bottles to more than $50000 for rare collectible bottles from closed stills
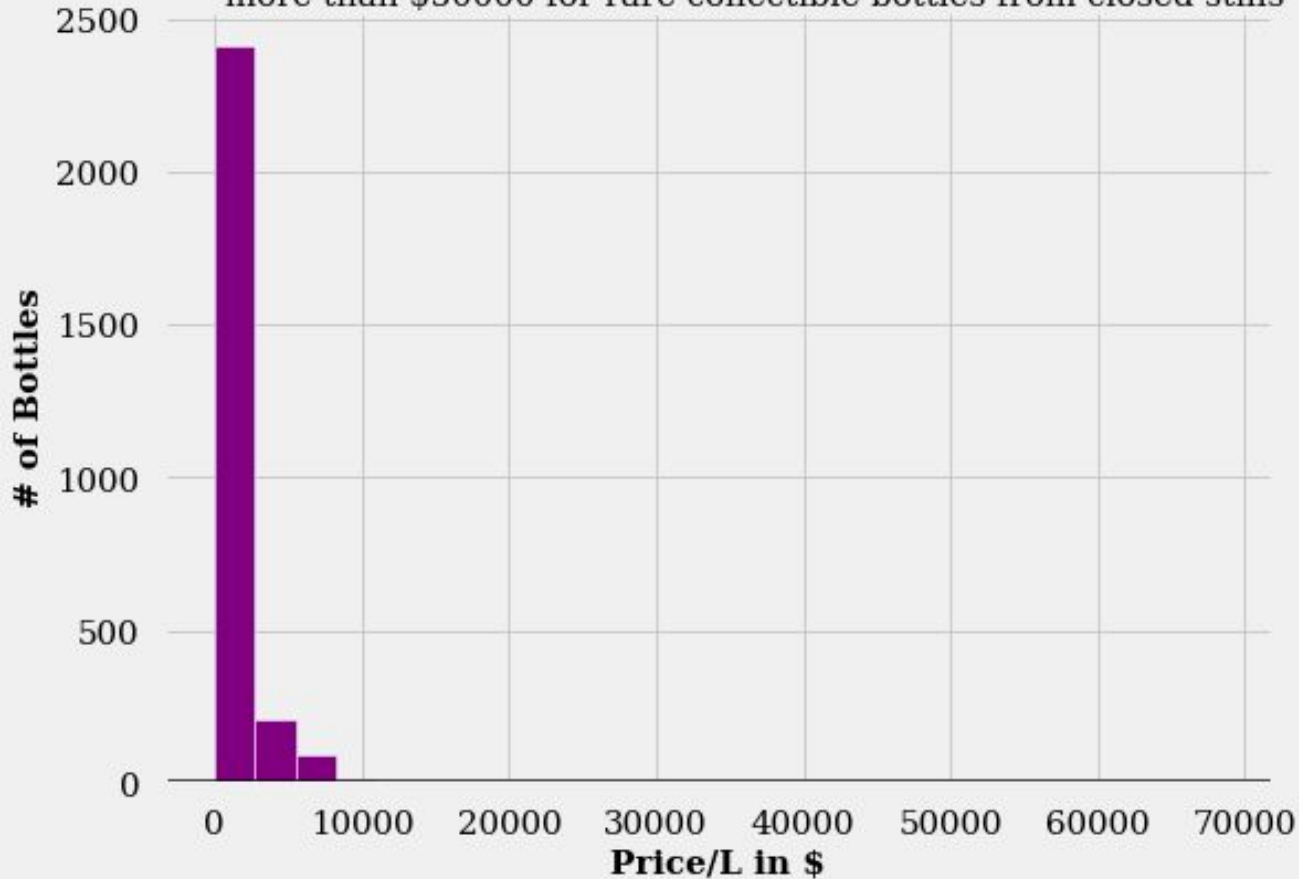
Whisky price range on TWE

Data Source : The WhiskyExchange | Anirudh Kashyap

Classification Problem

- DiSCUS classification (700ml)

  - Value (Class 0)   -   < $50
  - High End (Class 1)   - $50 - $100
  - Premium (Class 2)   - $100 - $1000
  - Ultra High End (Class 3)   -   >$1000

Categorizing Whisky prices using DiSCUS classification

more than $100
less than $1000

more than $50
less than $100

more than $1000

under $50

# of Bottles

1200

1000

800

600

400

200

0

0   1   2   3

Price/L in $

Data Source : The WhiskyExchange | Anirudh Kashyap

Class distribution using DiSCUS grouping

Model Evaluation (DiSCUS)

logit = Logistic Regression()

- Interpretability

- Easy to understand results

- Direct information for TWE to apply

- Speed

logit = Logistic Regression()

- GridSearch CV + Logistic Regression

- Accuracy: 0.69

- Sensitivity: 0.77

Analyzing the factors that place whisky in class 0

To read the chart: Whisky is (y-value) times as likely to be in class 0 (<$50) if it contains (x-value)

Odd times as likely (y-axis): 0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00

Factors (x-axis): Status, light, 12, eight, blend, BottlingType, start, d_glen moray, origin, la, 2016, cocktail, 2017, 10yo, product

Data Source : The WhiskyExchange | Anirudh Kashyap
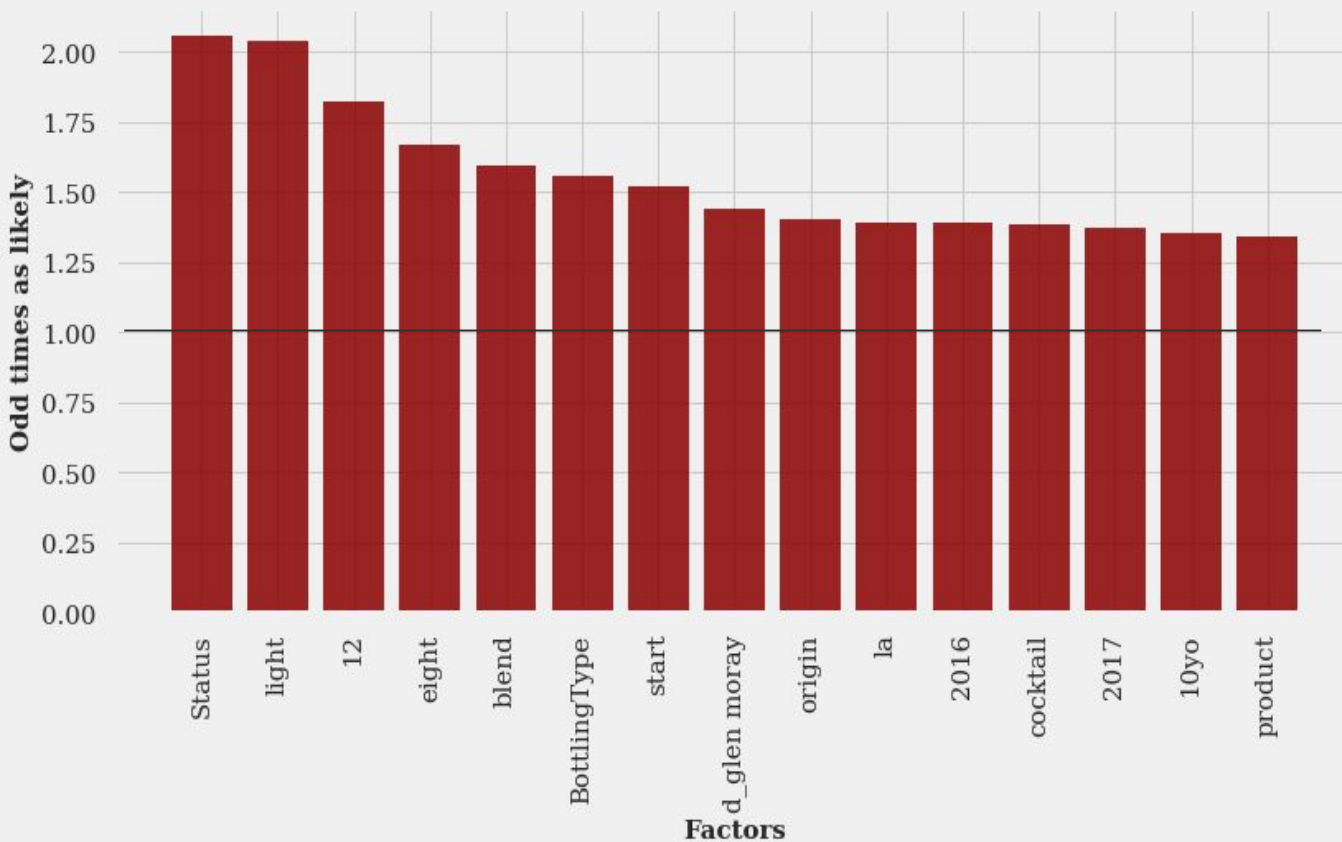
What are the factors (<$50)?

Status of distillery (Closed = 0)

Characteristics

BottlingType

Vintage

Age

Interpreting the chart (For TWE)

- If the distillery is Open, the whisky is 2.0 times as likely to be in Class 0 Vs if the distillery is Closed
- If the description contains the words 'light', 'blend', '10yo', the whisky is (Y-value) times as likely to be <$50

Analyzing the factors that place whisky in class 3

To read the chart: Whisky is (y-value) times as likely to be in class 3 (>$1000) if it contains (x-value)

Odd times as likely

Factors: VintageInfo, d_brora, r_islay, legendari, d_port ellen, r_campbeltown, w_single malt scotch, d_glen grant, 50, vintag, LimitedEditions, d_ardbeg, d_glenfiddich, sherri, famili

Data Source : The WhiskyExchange | Anirudh Kashyap
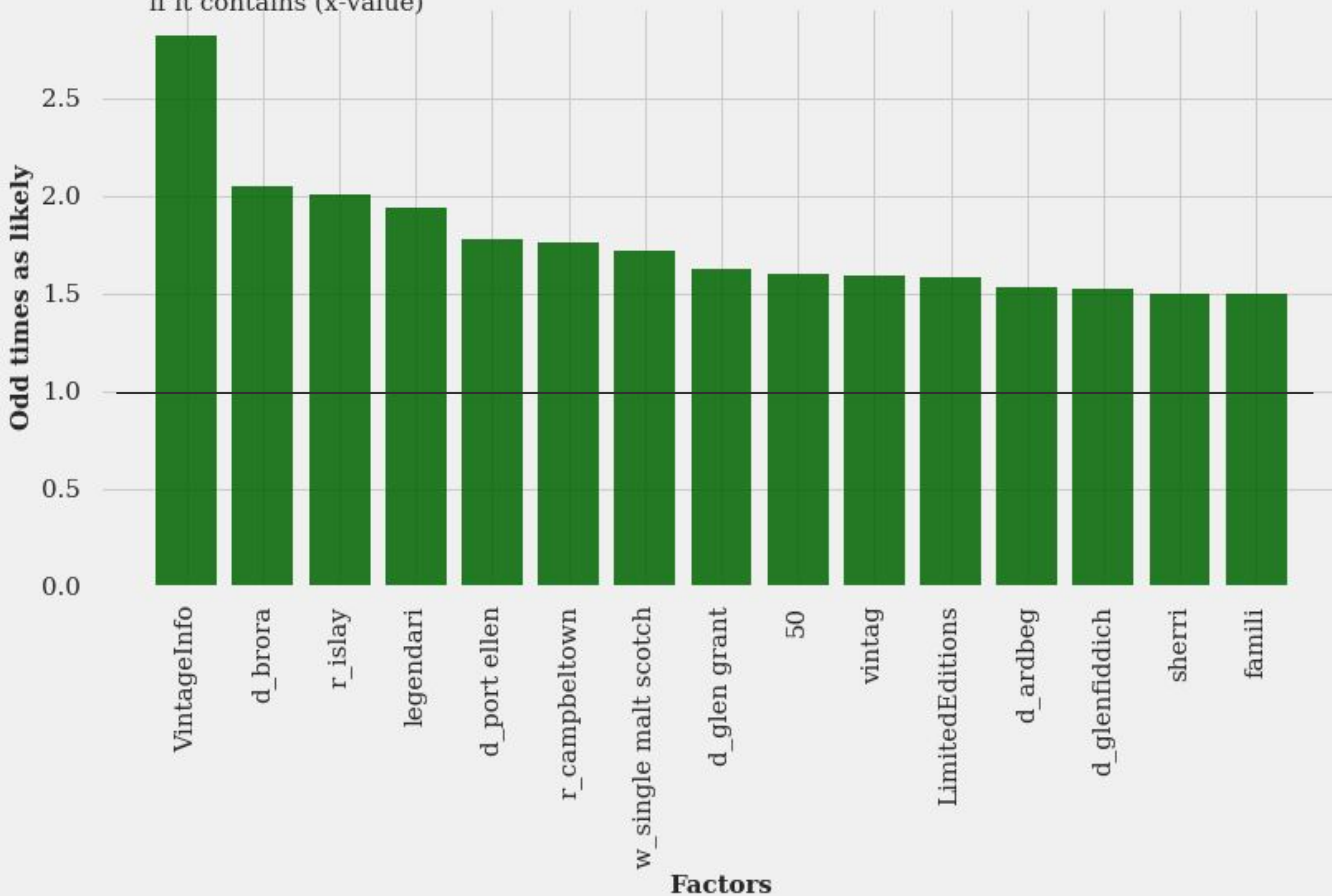
What are the factors (>$1000) ?

Vintage

Status

Type of whisky

Age
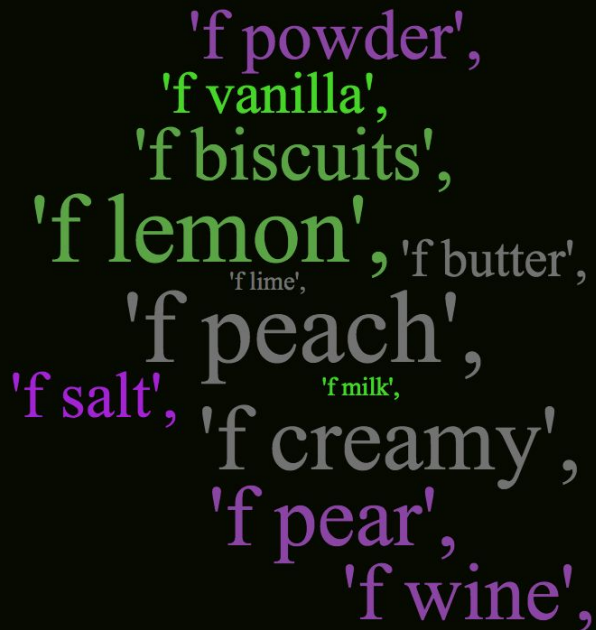
Editions

Interpreting the chart (For TWE)

- If the Vintage info is on the bottle/description, it is 2.5 times as likely to be in Class 3 (>$1000). Of course there are many other factors as well and this effect is compounded with various other predictors
- If the whisky is from brora, port ellen, it is expensive (duh!)
- Class 3 bottles have a higher chance of being limited editions, single malts, have the word 'legendary' in description

mmm .. f_flavors
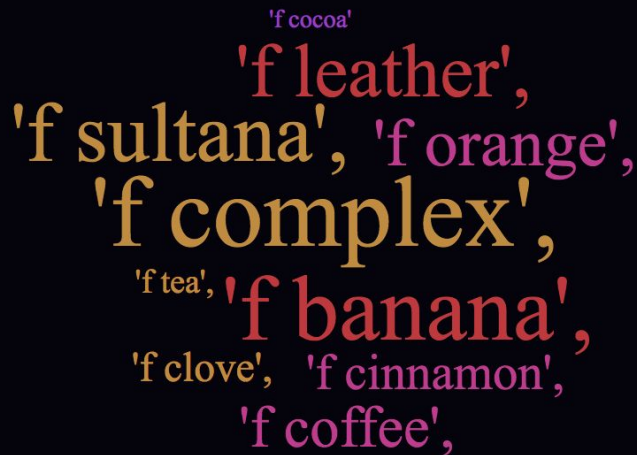
Class 0 (<$50)

'f powder',
'f vanilla',
'f biscuits',
'f lemon', 'f butter',
'f lime',
'f peach',
'f salt', 'f milk',
'f creamy',
'f pear',
'f wine',

Class 3 (>$1000)

'f cocoa'
'f leather',
'f sultana', 'f orange',
'f complex',
'f tea', 'f banana',
'f clove', 'f cinnamon',
'f coffee',

It is a numbers game

Class 0

12	2016	2017	10yo	1990s	100
31	2013	2011	21st	eight	seven

Class 1

15   10   2017   2015   11   1997   12yo

Class 2

1985   1984   1995   1970s   14
992   1980s

Class 3

1974   1954   1938   1941
1966   19yo   50yo

Take a guess?

- Status of distillery (Open = 1)

- Description (flavor, age)

An eight-year-old whisky from one of Diageo's lesser-known distilleries, Inchgower. Aged in an oloroso-sherry butt, this has notes of green herbs, vanilla and mint.

- Single Malt Scotch, 2016

- BottlingType - Independent

Actual Price

$48.72

Improving model performance

- GridSearchCV + RandomForest

- Accuracy & Precision: 0.7 for DiSCUS classes

- Drawbacks:
  - Only know feature importances

Model Evaluation
(Anirudh's classes)

Alternative classification

- Anirudh's classification (700ml)

  - Affordable   -   < $500
  - Are you crazy?!?   -   > $500
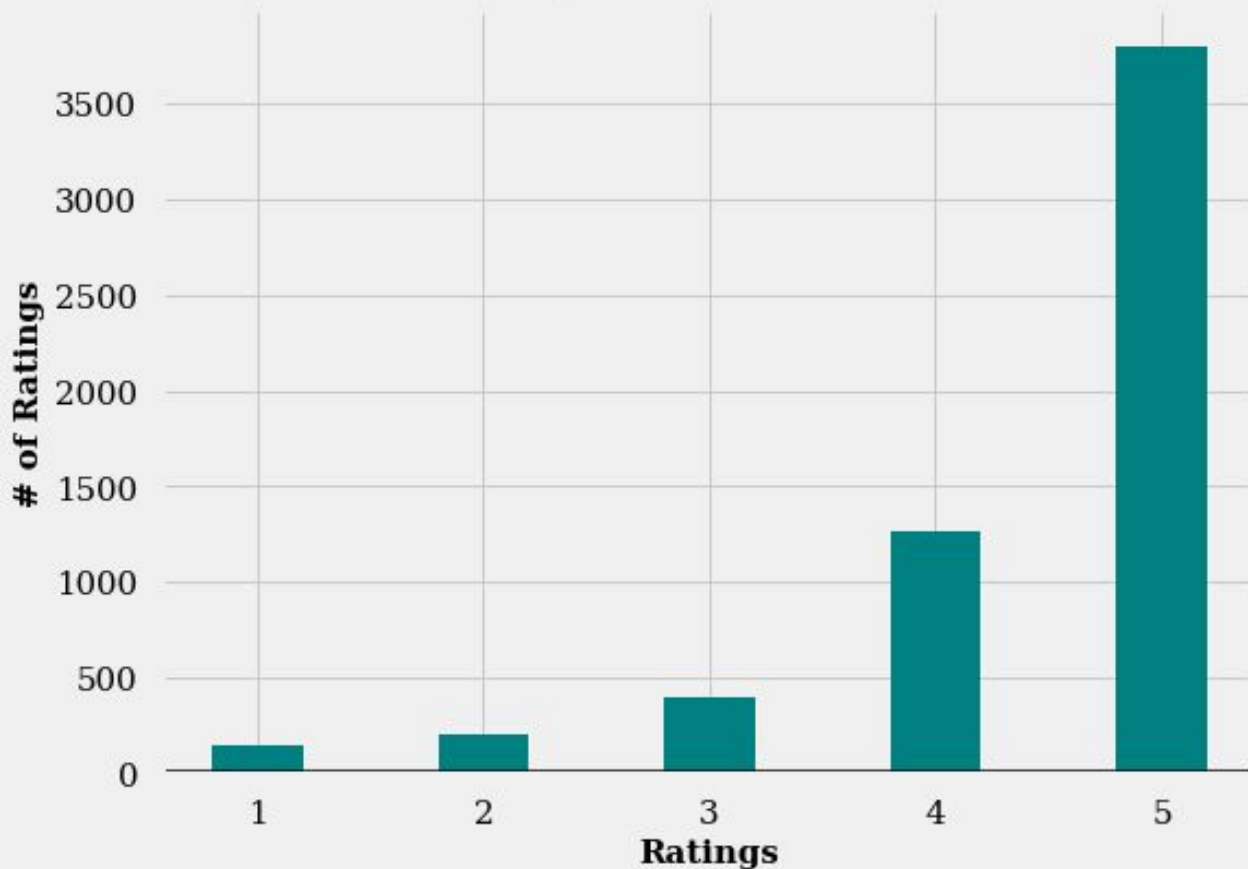
- Balanced classes

Improvement in scores

- Accuracy: 0.88
- Sensitivity: 0.85

- Similar scores using RandomForest

- >$500 keywords - oldest, incred, 1966, vintage, 50
  <$500 keywords - 2016, 2017, 10yo, refill, official

What are the customers saying?

Recommendation

- Case for a Like/Dislike system?
  - People have very different opinions of a 1-5 rating system
  - Netflix/Youtube recently switched to a thumbs up/down system
  - Better predictions with a binary classification system

We will compare 5.0 reviews Vs not 5.0 reviews

Review they wrote...
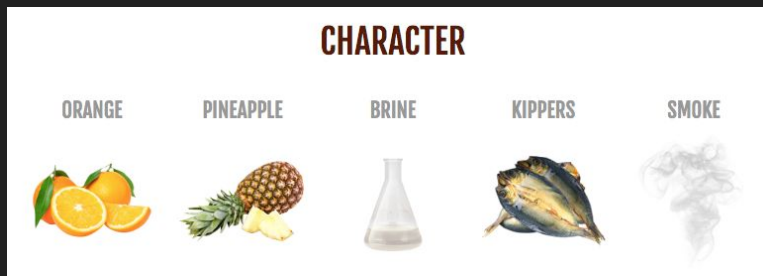
148	137	11
Flavor list created from Character Box	Flavors mentioned in comments	Flavors not described in reviews

CHARACTER

ORANGE   PINEAPPLE   BRINE   KIPPERS   SMOKE

Flavors in Character Box but not in Reviews
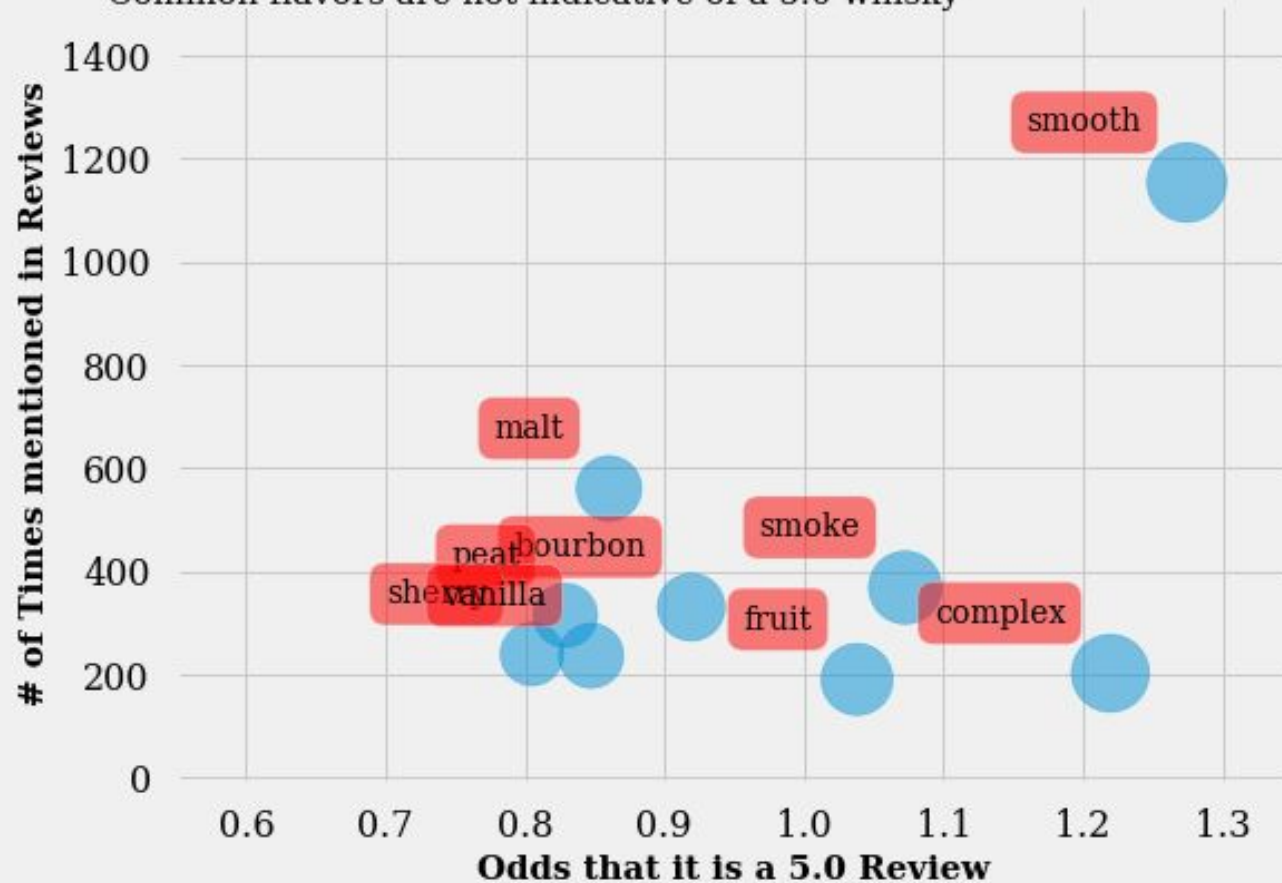
Pear Drops

Sultana

Caraway

Rosemary

Praline, Herb

Blackberry

Seashell, Matchbox

Top 10 Flavors listed Vs Odds its a 5.0 Review

Looking @ the Top 10 profiles identified by reviewers
Common flavors are not indicative of a 5.0 whisky

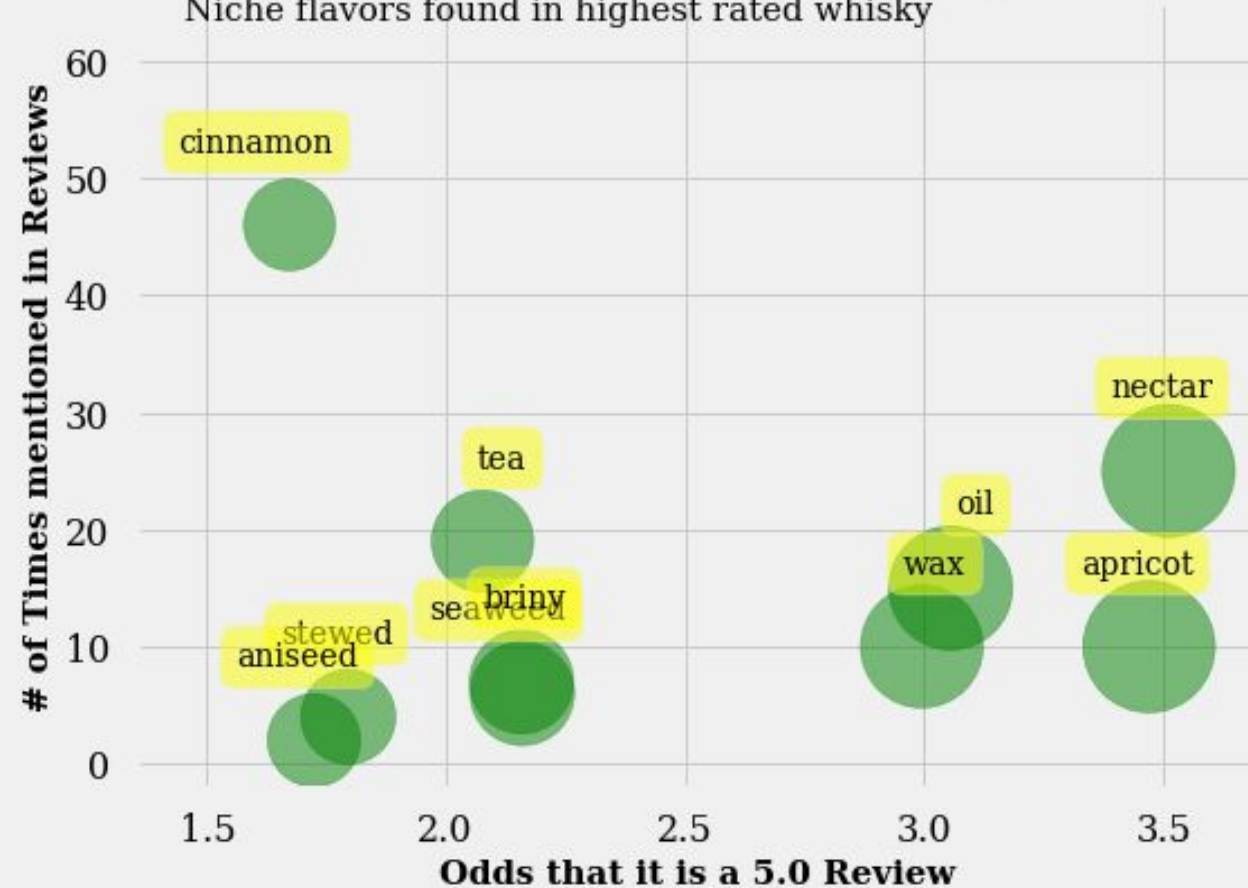Data Source : The WhiskyExchange | Anirudh Kashyap

Most whisky drinkers identify a whisky by the base flavors:

Smoothness, Vanilla, Peat, Bourbon, Sherry, light & Fruits

These flavors are dominant in a whisky profile

10 flavors in higest rated whisky

The flavors found in whisky with odds of having 5.0
Niche flavors found in highest rated whisky

# of Times mentioned in Reviews (y-axis, 0 to 60)

Odds that it is a 5.0 Review (x-axis, 1.5 to 3.5)

Labels on chart: cinnamon, tea, nectar, oil, wax, apricot, briny, seaweed, stewed, aniseed

Data Source : The WhiskyExchange | Anirudh Kashyap

Looking at whisky with 5.0 ratings we see flavor mentions:

Tea, Oil, Apricot, Cinnamon, Aniseed

Usually it takes a trained palate to discover these subtle flavors & not too many people can find them.
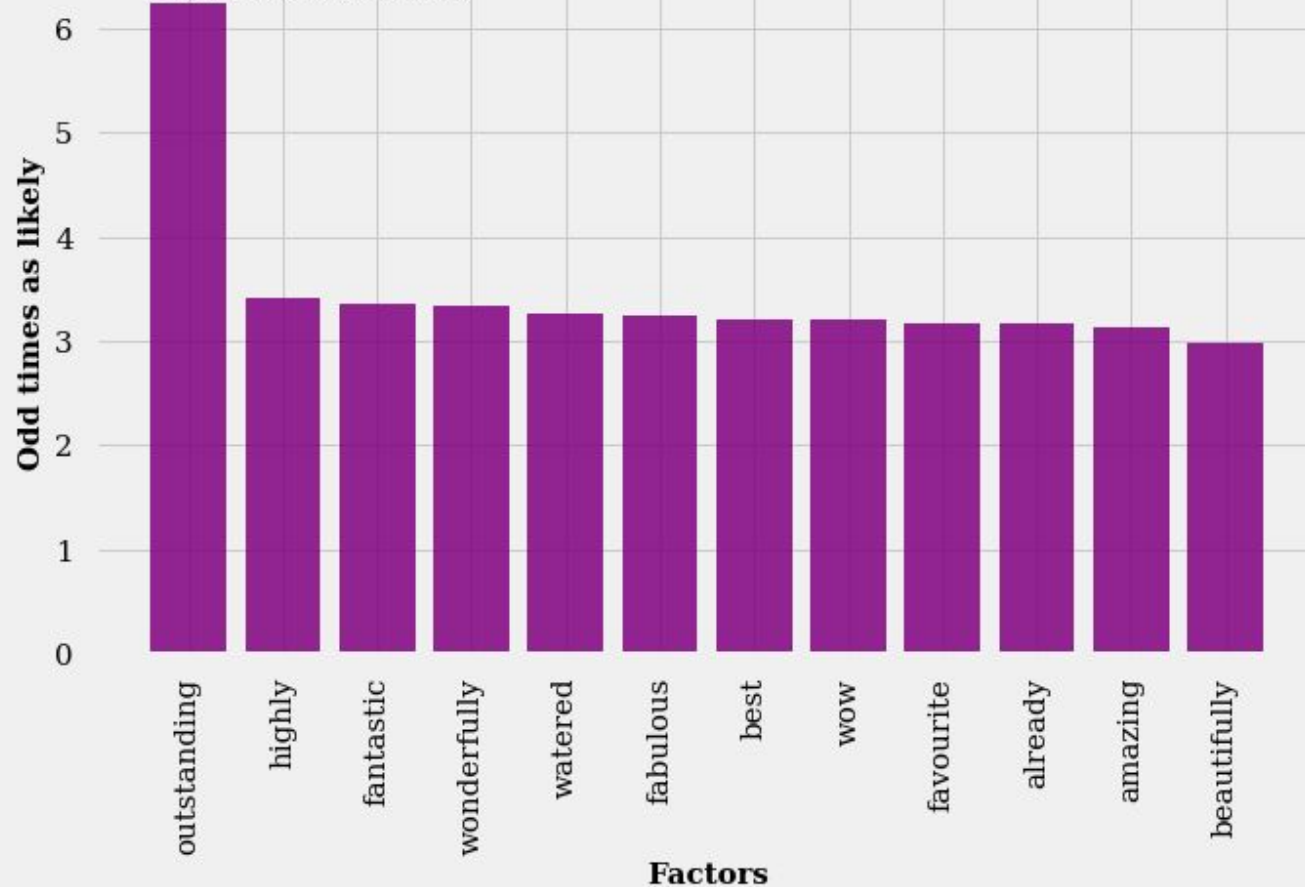
Keyword mention

- Flavors are not good predictors of ratings. Popular flavors are found in both highly reviewed & poorly reviewed whisky
- Most people can tell if whisky has smooth/peaty/sherry finish/vanilla notes
- But most can't identify heather/brine in a whisky. But those who do - give it high ratings

How are people describing a 5.0 rated whisky?

What words do reviewers use to describe a 5.0 whisky?

To read the chart: Whisky is (y-value) times as likely to have a 5.0 review if it contains (x-value)

Odd times as likely

Factors: outstanding, highly, fantastic, wonderfully, watered, fabulous, best, wow, favourite, already, amazing, beautifully

Data Source : The WhiskyExchange | Anirudh Kashyap

If the review contains the word 'Outstanding', it is 6 times as likely that the reviewer has left a 5.0 rating for the whisky

What are customers saying about < 5.0 whisky?

- Emotions
  Bad | Poor | Hit | Worst | Disappointing | Harsh

- Most common word
  - Ok

- 2-word pairs, specific words can be filtered as well.

Conclusions

Buy smart. Research much.

- The factors involved in purchasing a bottle of whisky (status, age, vintage, type, flavors)

- Watch out for pricey No Age statements with no age mentions

- Description has a ton of information

- Look beyond the packaging

Customer Feedback

- Emotions like outstanding, great, excellent are indicators of a 5.0 star whisky

- Emotions like ok, poor, watered are indicators of a non-5.0 star whisky

- Flavor reviews help identify whisky quality

Thank you,

Matt Brems, Matt Speck, Joe Klein
+
Class of DSI-6
+
The Whisky Exchange