# EH MACS?

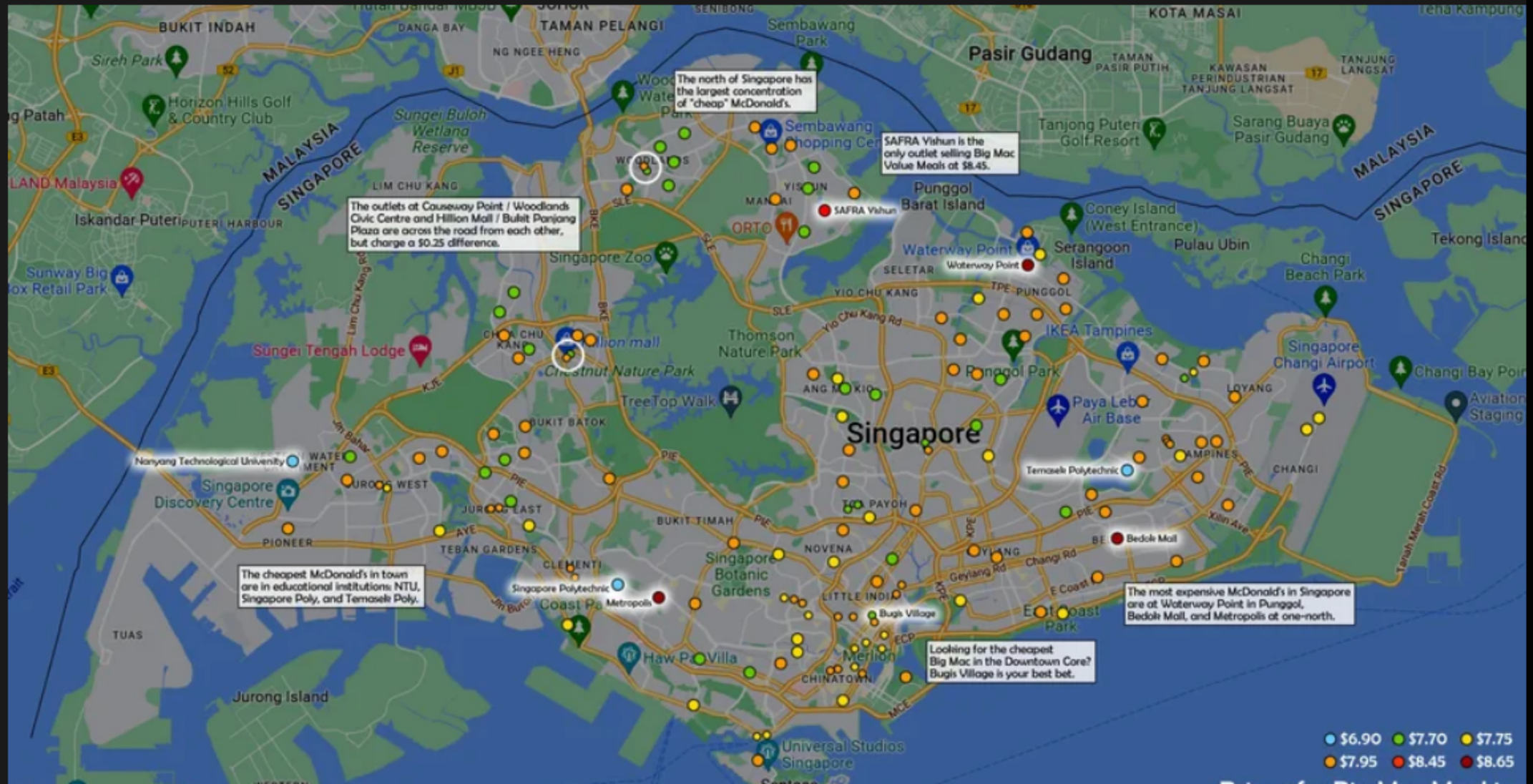
*A way too detailed way to answer a simple question*

Posted by u/EconomicSanction 3 months ago 6 5 4

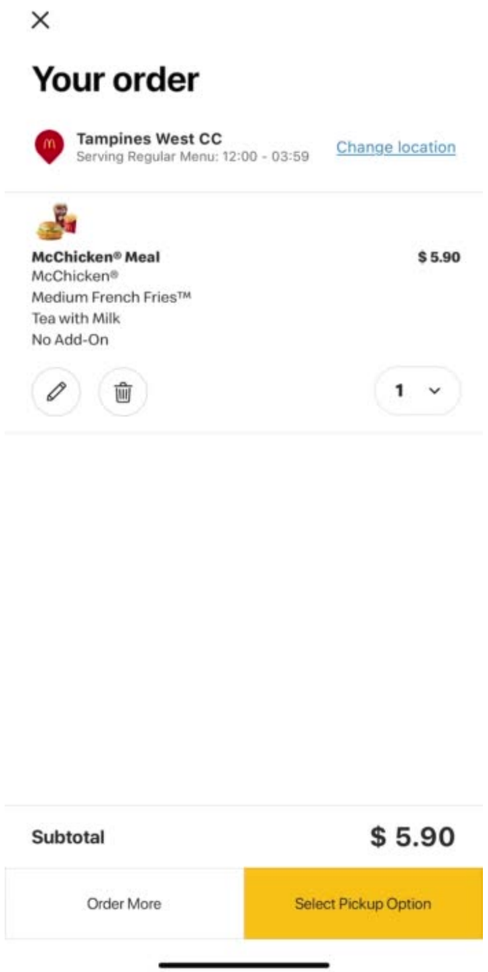
2.6k

# Price of a Big Mac Meal across Singapore

I Made This



## Manual Scraping




## Web Scraping

Octoparse

Google Maps

=

**Supply Factors**

Ingredients

Salary

Rental

+

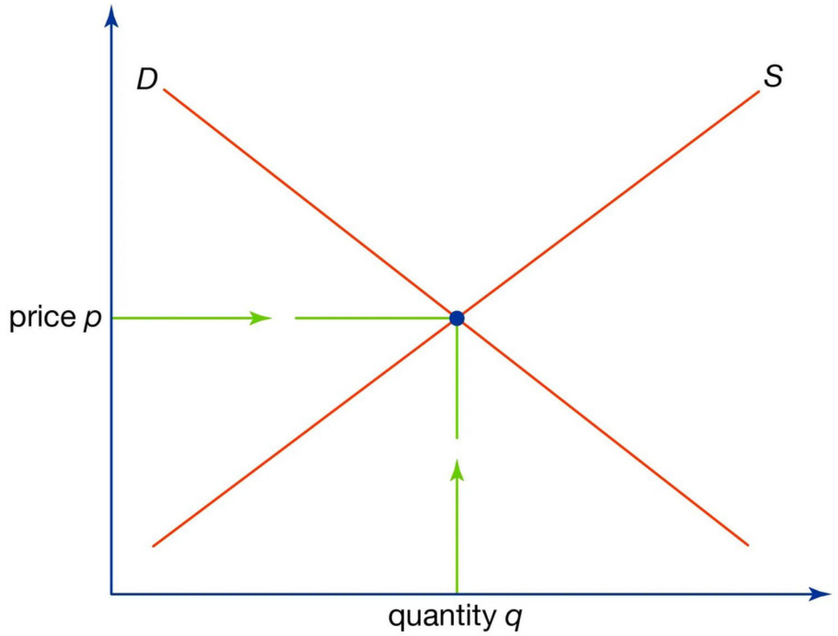
**Demand Factors**

# of Customers

Income of Customers

Branding

**Supply and demand**



**Supply Factors**
Ingredients
Salary
Rental

**Demand Factors**
# of Customers
Income of Customers
Branding

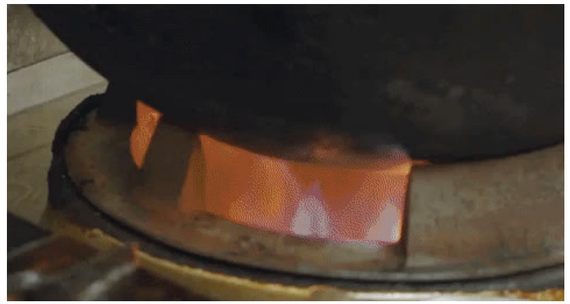
**Why McDonald's? Because they are the closest to being competitively priced!**

- Menu actually has a price difference depending on where you buy it
- Huge presence and good coverage of the island to test hypothesis
- 140+ stores operating since 1979
- Every restaurant is "the same" so we have less factors to consider when solving our equation

**Supply Factors**
Ingredients
Salary
Rental

**Demand Factors**
# of Customers
Income of Customers
Branding

**After that we can solve every other F&B to help capture market inefficiencies!**

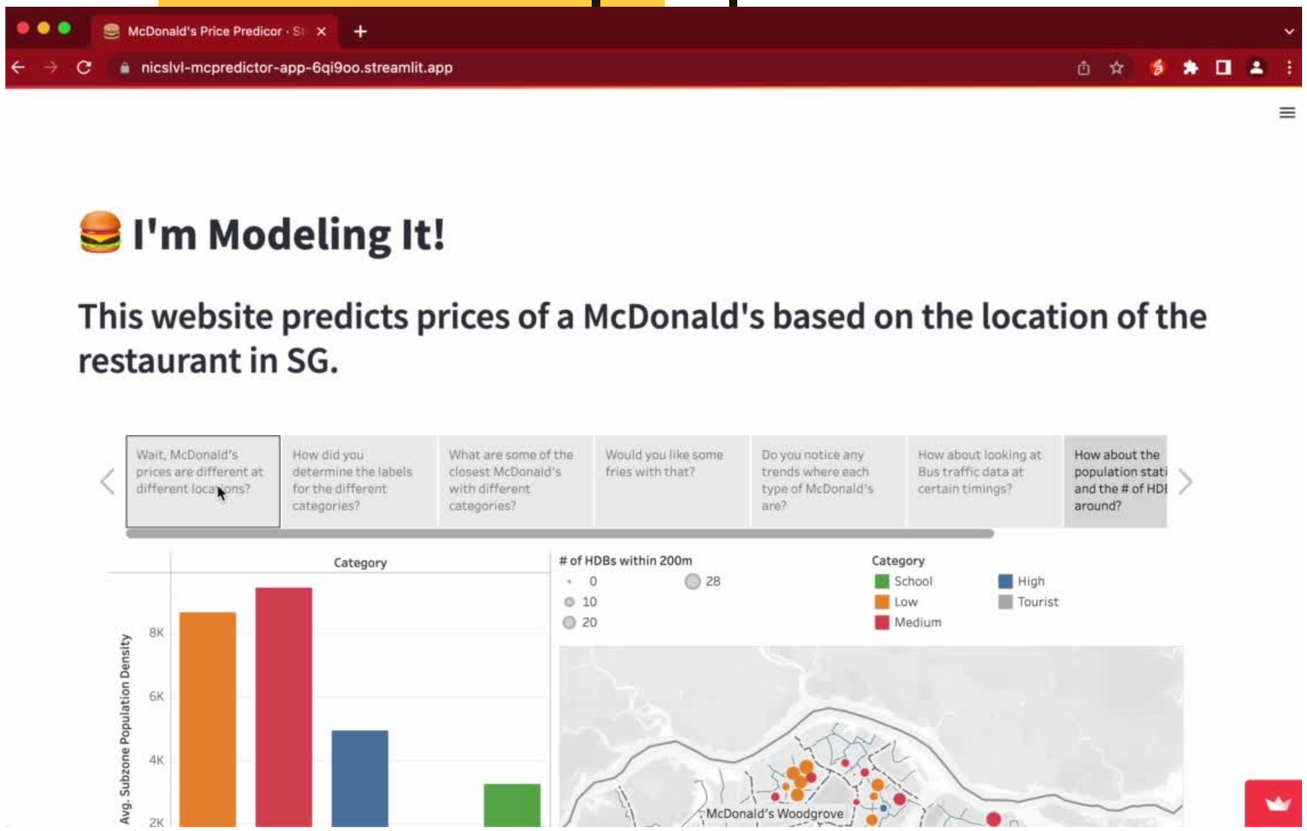


**Supply Factors**
Ingredients
Salary
Rental

**Demand Factors**
# of Customers
Income of Customers
Branding

# Demo!

SCAN ME


https://qrco.de/bdqqbr

# Feature Selection

| **An approximation of...** | **using...** | **by calculating...** | **leaving us with...** |
|---|---|---|---|
| # of Customers (variable) | Bus Traffic | Within X distance*, how much (in/out) bus traffic was there based on 24 hours on a (weekend/weekday) around a Macs? | 24 hours x 2 day types x 2 traffic types x 4 distances = **384 features** |
| Rental, Income of Customers | HDB Count | Within X distance*, how many hdbs are there around a Macs? | 4 distances = **4 features** |
| # of Customers (fixed), Income of Customers | Subzone Data | For a McDonald's what is the<br>• Subzone Population?<br>• Subzone Population Density?<br>• Subzone Income?<br>• Subzone Average Income? | **4 features** |

** There were 4 distances selected: 200m, 500m, 1km, 2km*

# 1. Bus Traffic

## ≈ # of Customers (variable)

- We assume a fixed % of customers will go to Macs
- Data taken from LTA, 5083 bus stops
- Only past 4 months data available (Nov 22 - Feb 23)
- # of People who tap in / out at a particular stop
  - Weekend or weekday
  - What hour of the day (0-23)

# 2. # of HDB

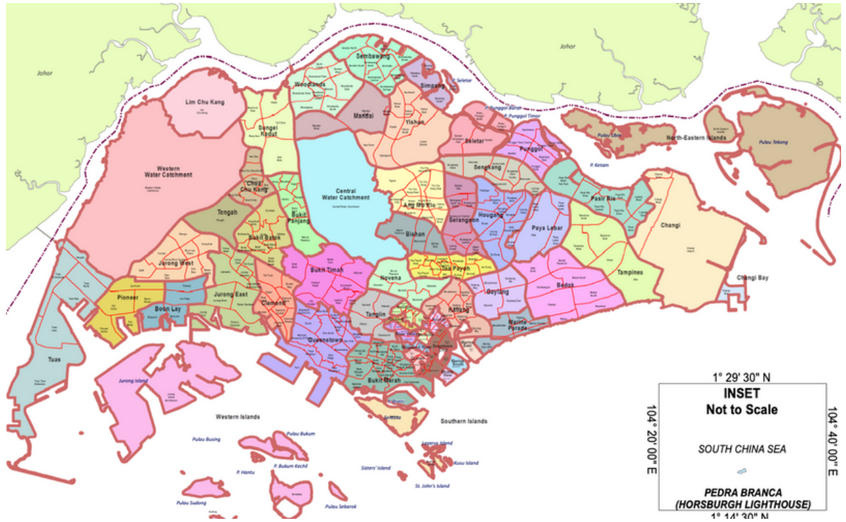
## ≈ Rental, Income of Customers

- Initially wanted Public:Private property ratio. Land is scarce in Singapore, if its not used for HDB, its used for private / commercial purposes
- Data taken from data.gov, 12,626 different HDB blocks
- Last updated 4 Jan 2023


# 3. Subzone Data

## ≈ # of Customers (fixed), Income of Customers

- We assume a fixed number of people living around the area to go to that McDonald's based on their flat type
- Data taken from Singstat, 332 different subzones
- Last updated Jun 2022



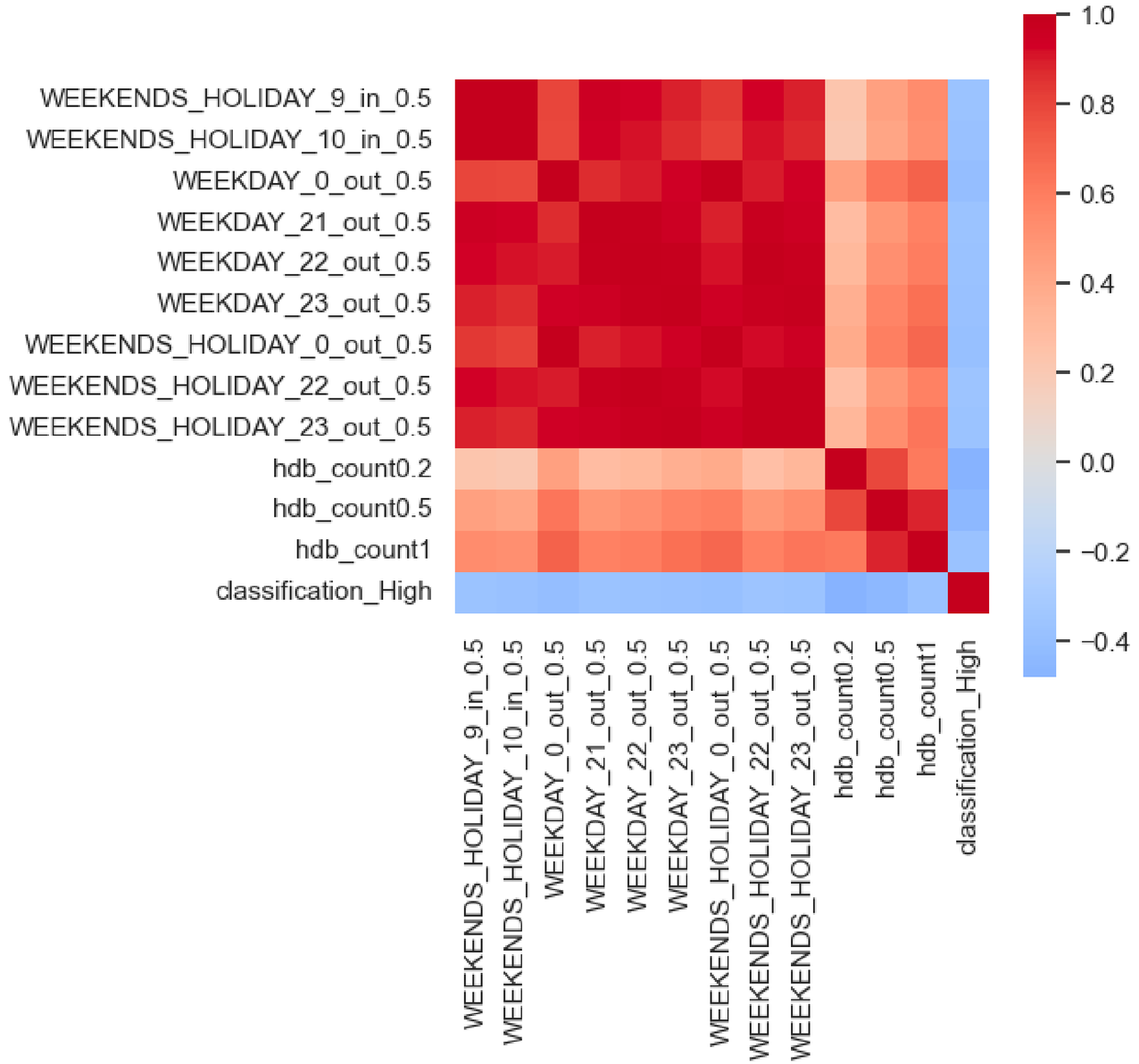

We look at features **>0.36 absolute correlation**

**01**

Tap in timings in the morning and tap out timings at night are negatively correlated

**02**

HDB count is negatively correlated

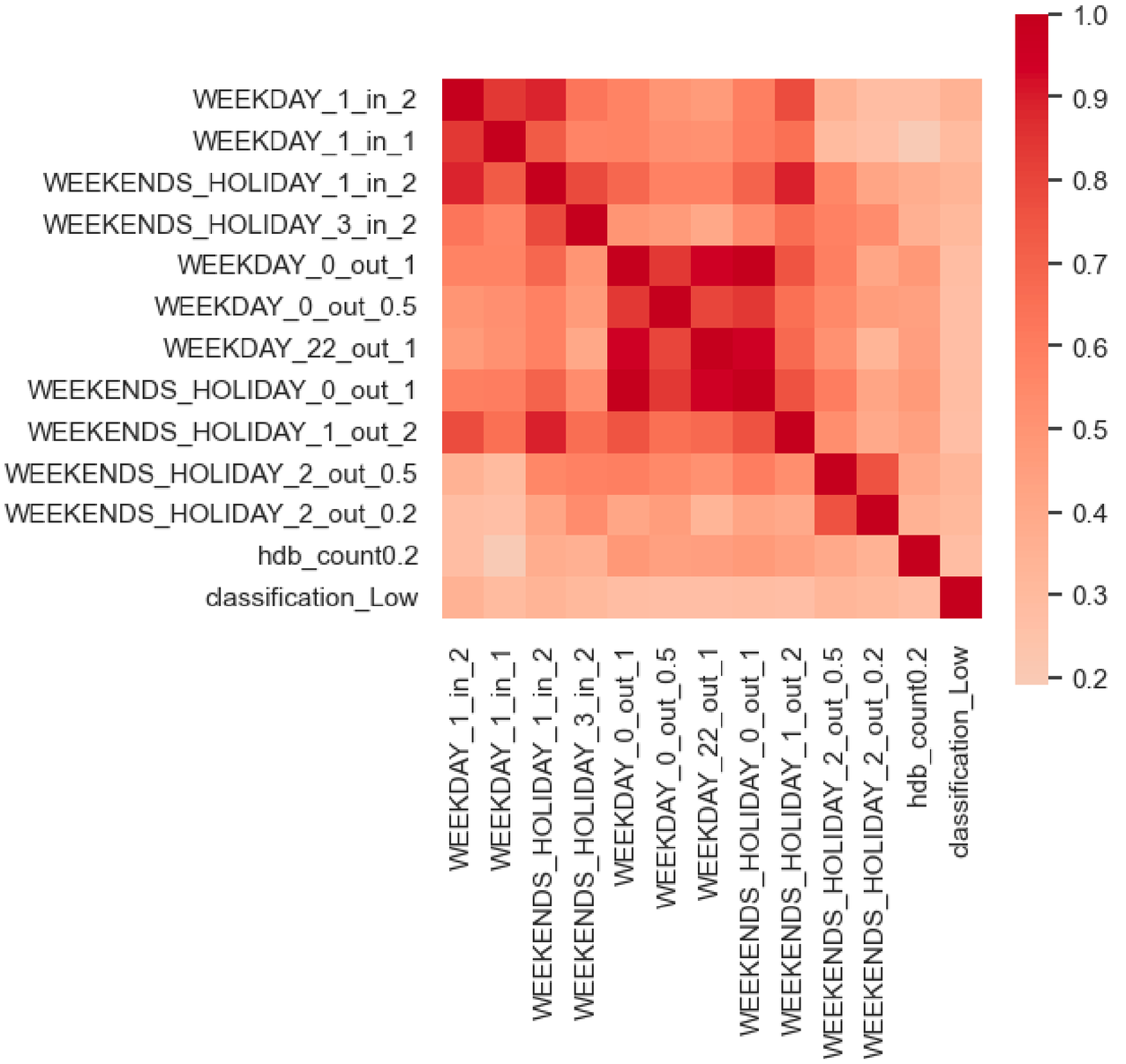



# "HIGH" McDonald's

We look at features **>0.27 absolute correlation**

# 01

Tap in timings in the early morning and tap out timings at night are positively correlated

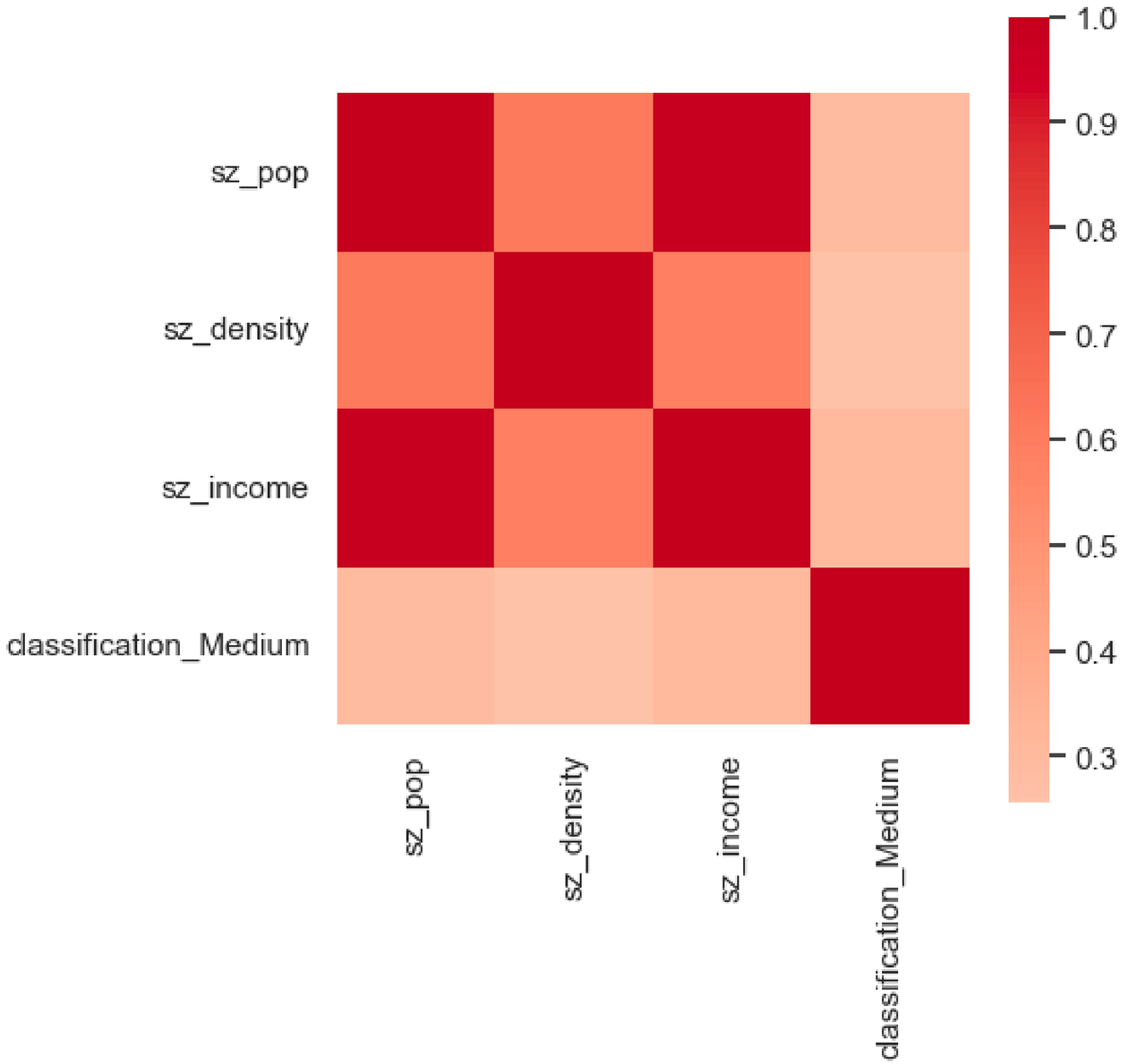



# "LOW" McDonald's

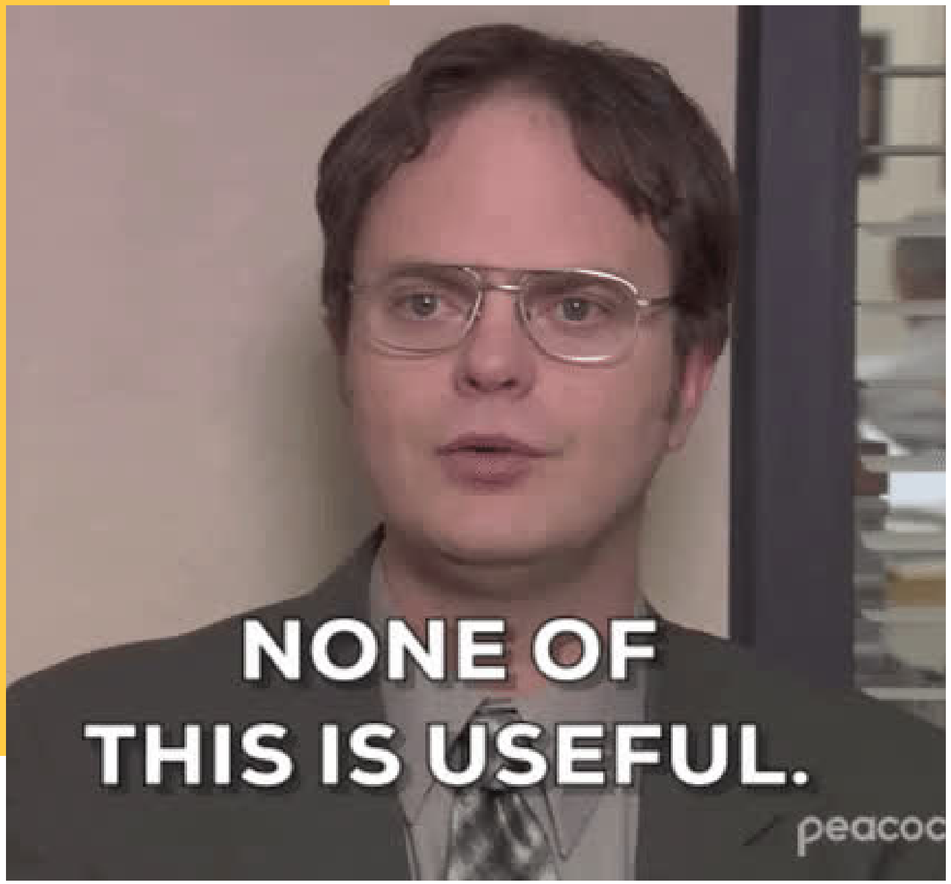
We look at features **>0.23 absolute correlation**

# 01

Somewhat positive correlation to subzone population, density and income




# "MED" McDonald's

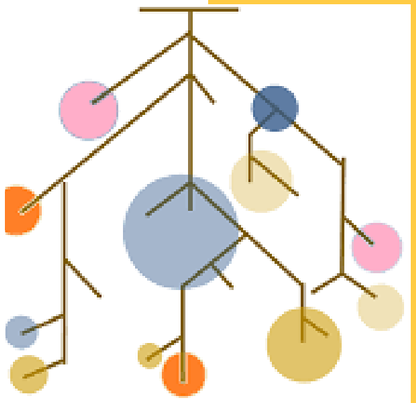



# Unused features

- Planning area Makan Index
- Hawker Centre count
- 24 hours Macs or not
- Average age of HDB in the subzone
- Subzone average income

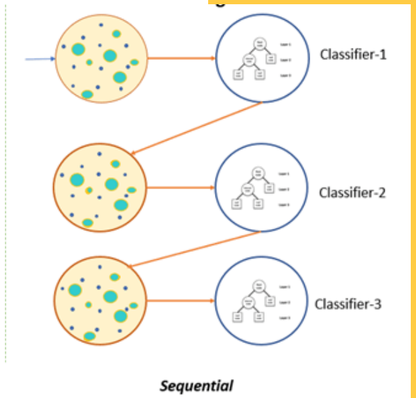
# Model Results

- Baseline is 33% to get a classification right
- All algorithms were oversampled to make up for the class imbalance
- We aim for accuracy because all classes are equally important to get right to test out the hypothesis



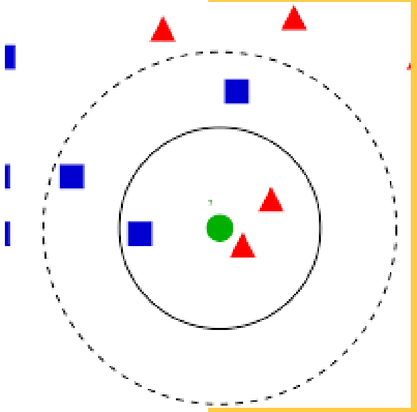
### Random Forest

Stratified Kfolds Train: 0.73
Test: 0.69
Split: 75 : 25




### Gradient Boosting Classifier

Stratified Kfolds Train: 0.57
Test: 0.57
Split: 75 : 25



### KNN

Stratified Kfolds Train: 0.6
Test: 0.5
Split: 75:25

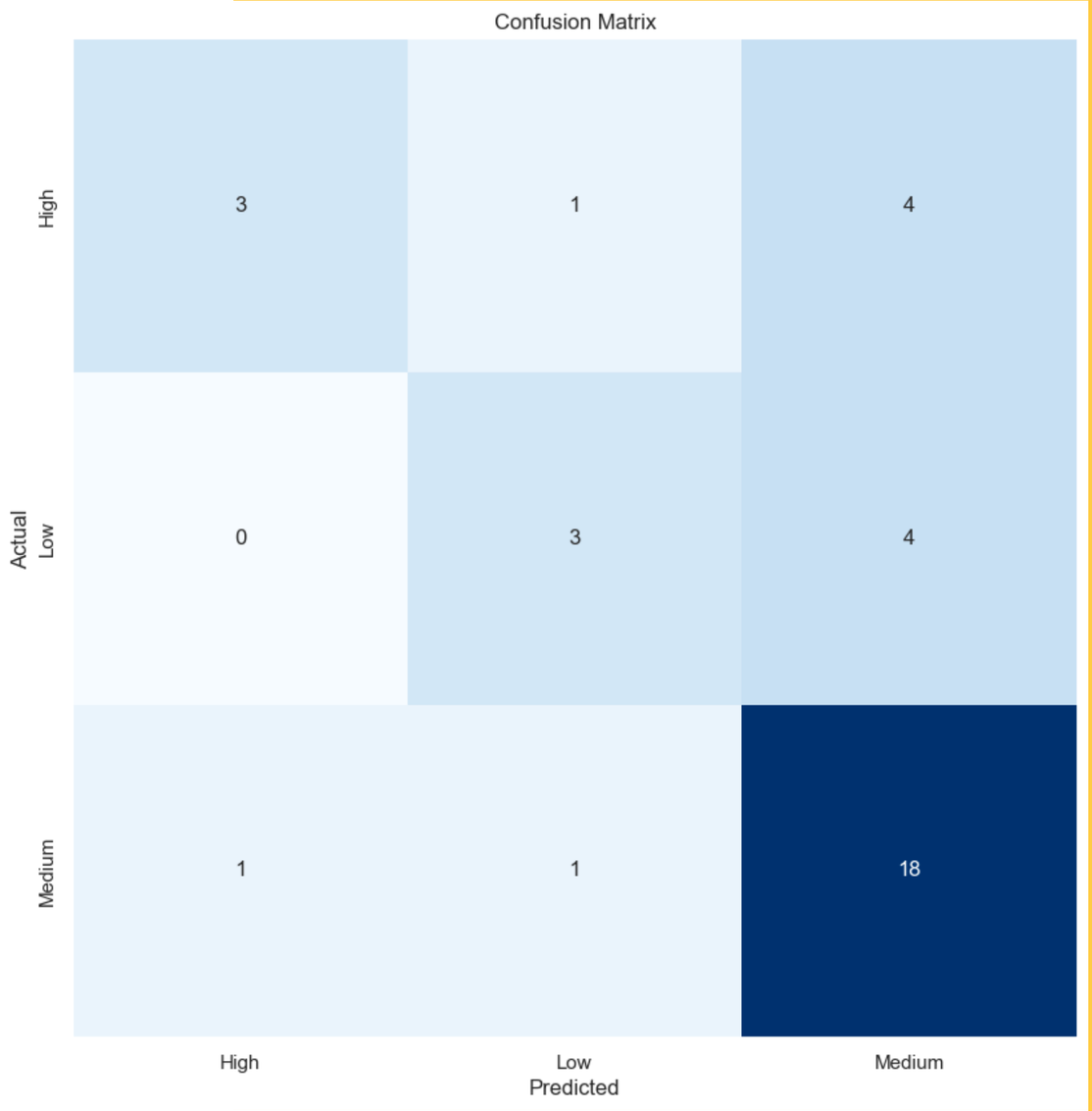
# Model Results

To improve past 69% **add more data!**

- New features are required to predict the medium "mistakes"
- Consider other F&B that have different pricing in their menus

**Medium but should be Low**
McDonald's Bukit Panjang Plaza
McDonald's Woodlands Civic Centre
McDonald's Khatib
McDonald's Yew Tee Point
**Medium but should be High**
Raffles City
McDonald's Toa Payoh SAFRA
McDonald's Shell Havelock
McDonald's Kallang (Stadium)

Others: McDonald's Keat Hong,  McDonald's Siglap, McDonald's Bishan Park



# THANK YOU!

Q & A

SCAN ME

**Demo:** https://qrco.de/bdqqbr

**Github:** https://github.com/NicsLvl/McPredictor