

# *From Bricks to Bucks*

## Exploring Influential Variables in Sales Patterns of Lego Bricks

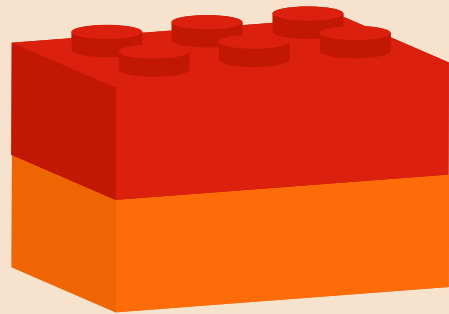
By Samuel Love



# Aims

**'Can we accurately predict the secondary prices of Lego sets using a defined set of characteristics?'**

'Do Lego sets based on licensed intellectual properties differ from Lego original sets?'



Trends in characteristics

Machine learning predictive models

# Objectives

Data visualisation

Predictive modelling

Time series

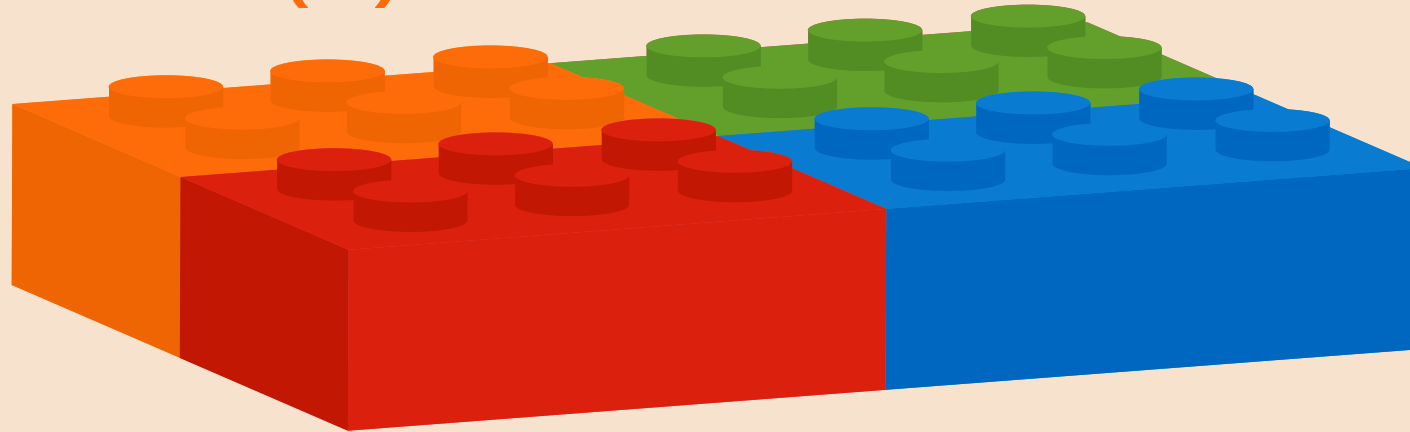
Cluster analysis

Feature ranking

# Data Source

Main Dataset (JB)

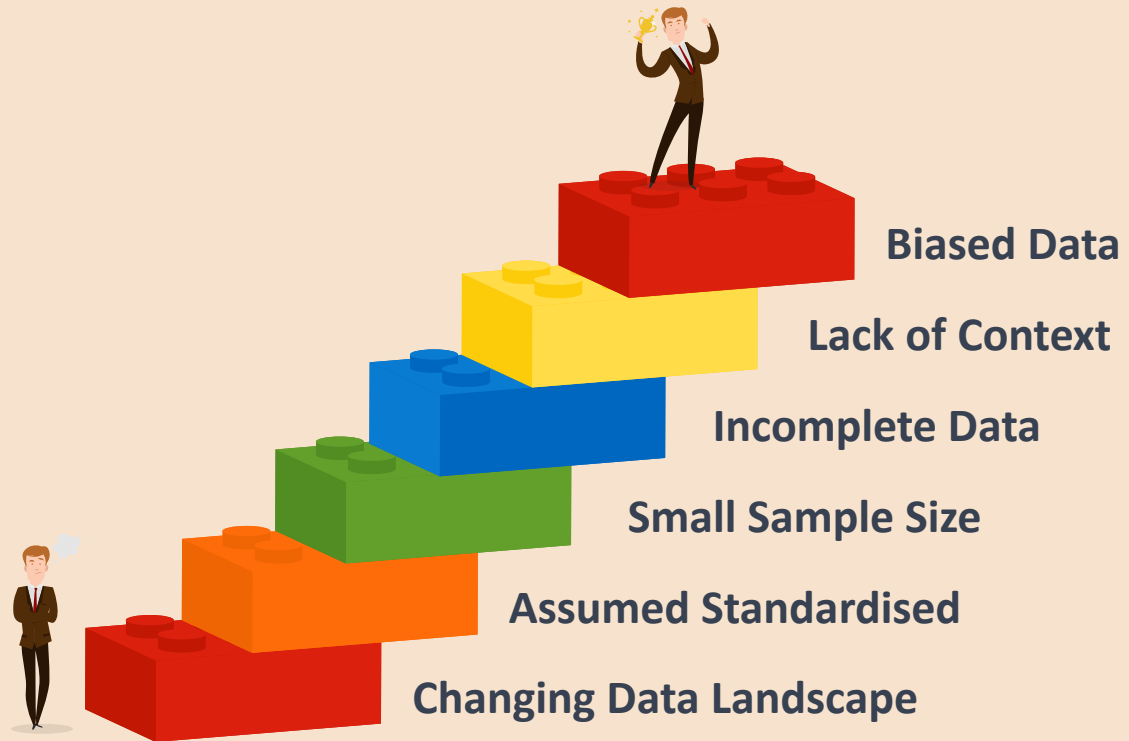
2022 Price Dataset (MP)



2015 Price Dataset (D15)

2018/2019 Price Dataset (D18)

# Data Limitations



# Method- Visualisations



## Step 1

Clean raw datasets

## Step 2

Create and clean 'Merged'

## Step 3

Generate clusters

## Step 4

Produce visualisations

# Method- Machine Learning

## Step 1

Create price subsets from 'Merged'

## Step 2

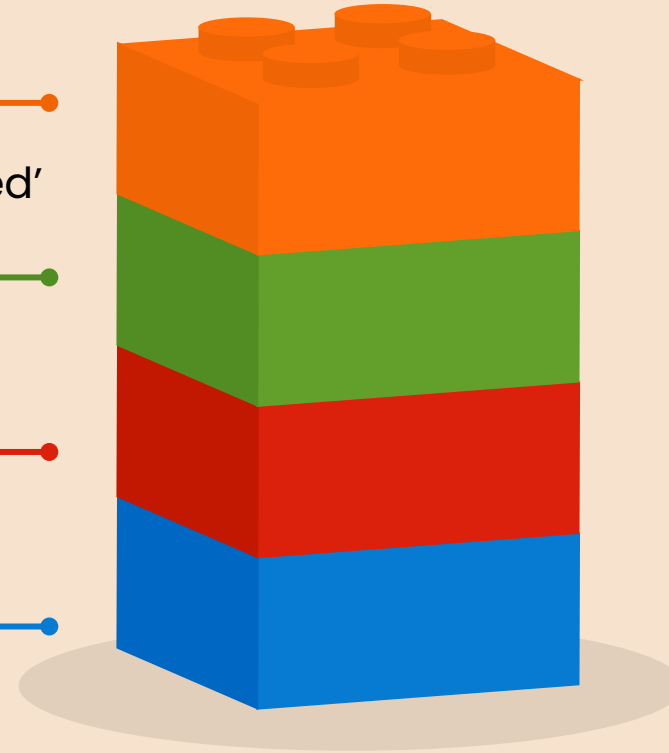
Train models

## Step 3

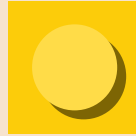
Visualise model metrics

## Step 4

Test best performing models



# Results – Visualisations



**Increasing number of sets**



**Characteristics diversifying**



**Average price variation**



**Price correlates with complexity**



**Organic clustering potential**



# Results – Predicting 2019 Prices



**Support Vector Machine best R2 and RMSE**



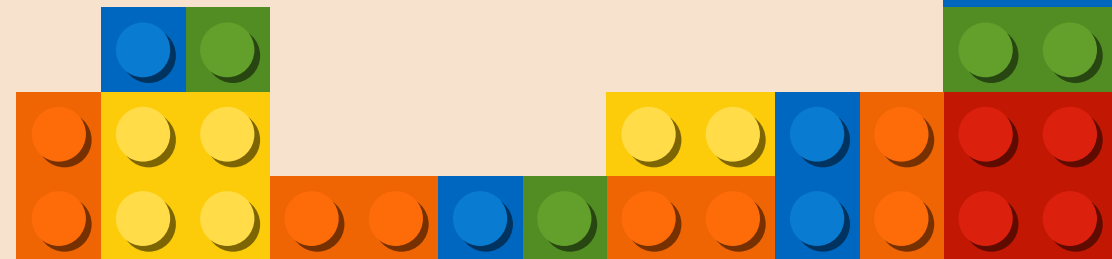
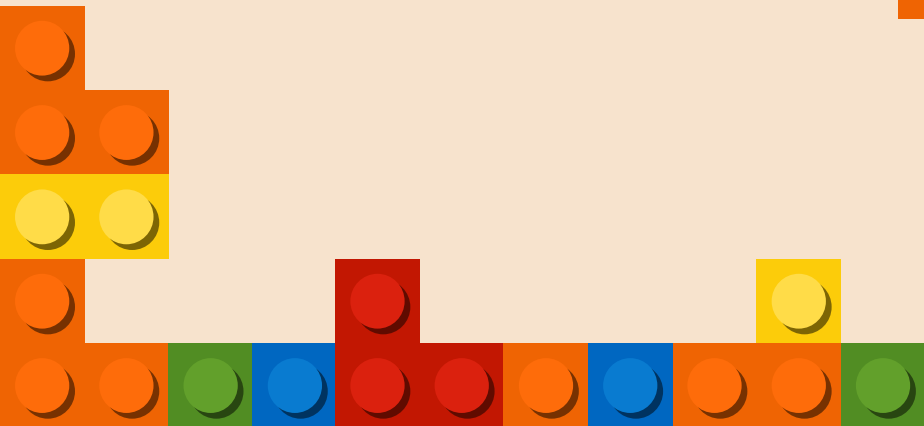
**Elastic Net best MAE**



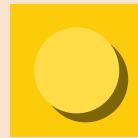
**Elastic Net top features:  
price2018, RRP, price2015, cluster, pieces**



**Random Forest top features:  
price2018, cluster, price2015, pieces, RRP**



# Results – Predicting 2022 Prices



**Elastic Net best R2 and RMSE**



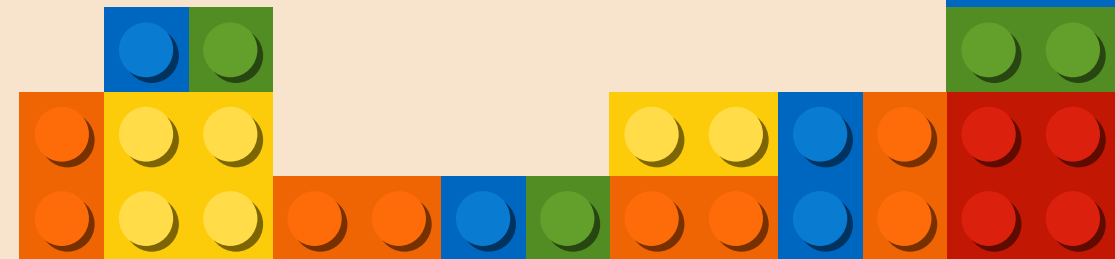
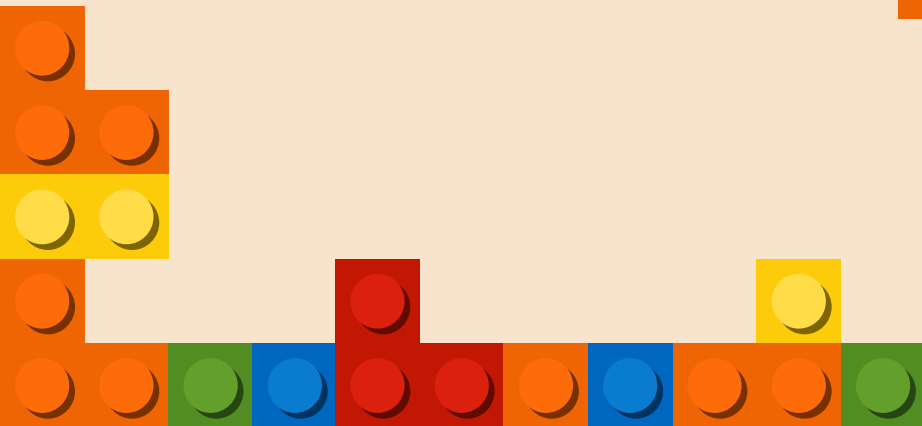
**Support Vector Machine best MAE**



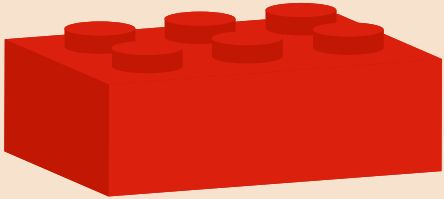
**Elastic Net top features:  
cluster, pieces, yearReleased**



**Random Forest top features:  
cluster, pieces, yearReleased**



# Conclusions



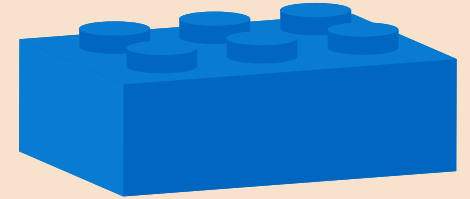
**Increasing  
complexity  
over time**



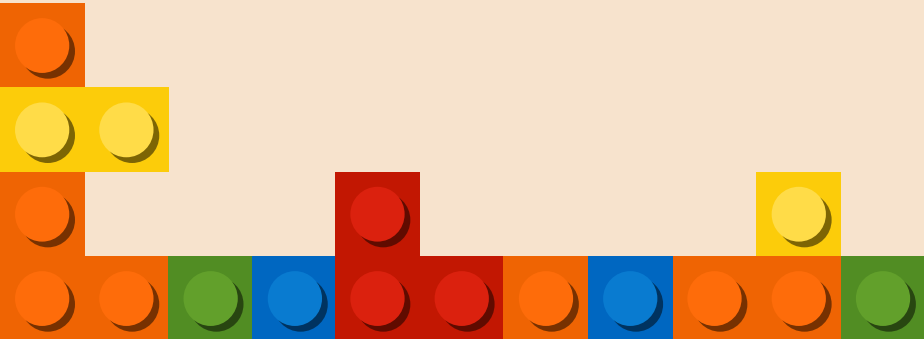
**High average  
price variation**



**Price variables  
best predictors**



**No obvious  
optimal model**



**Thank you for listening.**

**A special thanks to**



**Paul Benden**



**Phil Davies**

Any Questions