Declining Birth Rate in Singapore: SC1015 Mini Project

B124 Team 9
Kenneth Tan
Audric Ho

Table of Contents

01
Problem Definition

Singapore

Singapore's total fertility rate drops to historic low of 1.05

Singapore's total fertility rate has been falling for many years. The previous low was 1.1 in 2020.

The Problem

What are the key factors contributing to the declining birthrate in Singapore? How can we address this issue and encourage more births in Singapore?

02
Data Collection and Cleaning

DEPARTMENT OF

STATISTICS SINGAPORE

Empowering You with Trusted Data

Cleaning the Data

- Drop rows with NaN Values
- Transpose the Dataframe
- Convert the Year column to Datetime format
- Extract columns
- Merge datasets

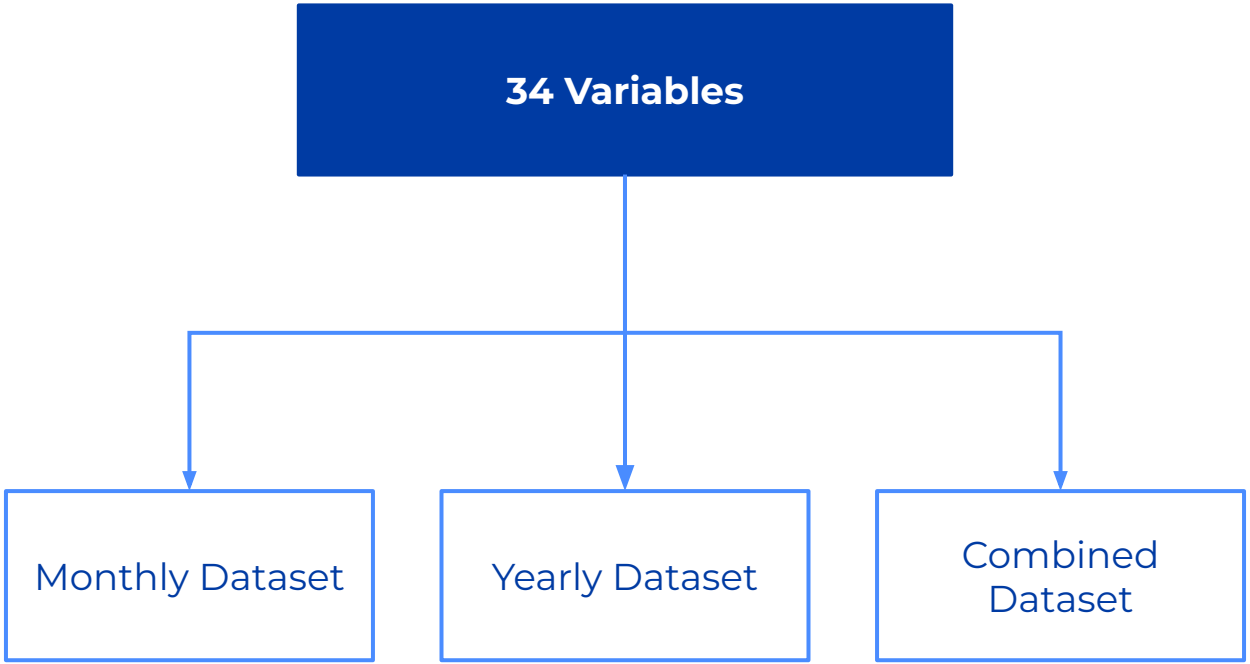
Theme: Population
Subject: Births and Fertility
Topic: Births, Crude Birth Rate (CBR), Total Fertility Rate (TFR) and Reproduction Rates
Table Title: Births And Fertility Rates, Annual

Data last updated: 24/02/2023
Source: SINGAPORE DEPARTMENT OF STATISTICS, IMMIGRATION AND CHECKPOINTS AUTHORITY

Data Series	2022	2021	2020	2019	2018
Total Fertility Rate (Per Female)	1.05	1.12	1.1	1.14	1.14
15 - 19 Years (Per Thousand Fe	2.1	2.2	2.3	2.5	2.5
20 - 24 Years (Per Thousand Fe	11.3	11.7	12.7	12.7	14.4
25 - 29 Years (Per Thousand Fe	49.1	53.4	54.6	59.4	60.6
30 - 34 Years (Per Thousand Fe	87.1	92.9	90.8	92.4	92.9
35 - 39 Years (Per Thousand Fe	49.5	53.6	49	50.1	48.4
40 - 44 Years (Per Thousand Fe	9.8	10.2	9.5	9.9	8.8
45 - 49 Years (Per Thousand Fe	0.4	0.3	0.5	0.4	0.5
Chinese (Per Female)	0.87	0.96	0.94	0.99	0.98
Malays (Per Female)	1.83	1.82	1.82	1.8	1.85
Indians (Per Female)	1.01	1.05	0.96	0.98	1
Gross Reproduction Rate (Per Fe	na	0.54	0.53	0.56	0.56
Net Reproduction Rate (Per Fem	na	0.54	0.53	0.56	0.55
Crude Birth Rate (Per Thousand	8	8.6	8.5	8.8	8.8
Total Live-Births (Number)	35,724	38,672	38,590	39,279	39,039
Resident Live-Births (Number)	32,417	34,183	34,233	35,330	35,040

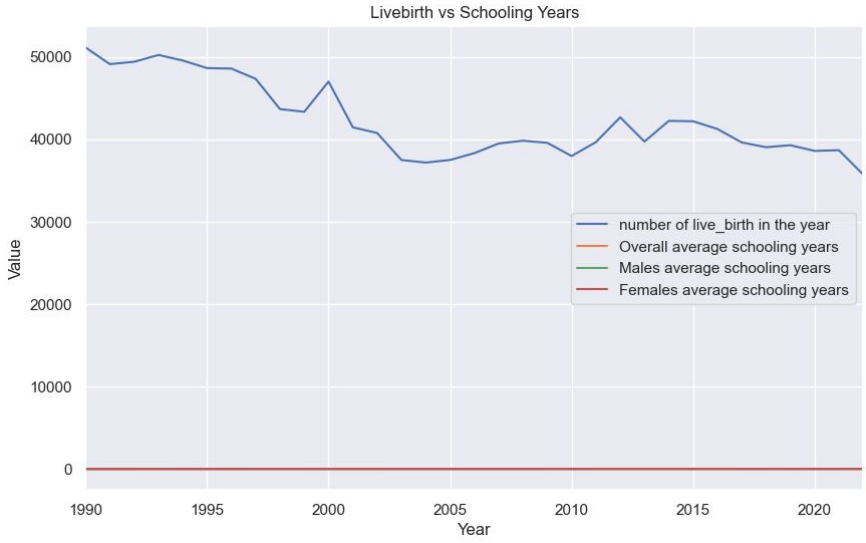
Footnotes:
Figures for 2022 are preliminary. For more information, please refer to the 'Population Trends' publication (www.singstat.gov.sg/publications/population/population-trends), the Infographics on 'To
Total Fertility Rate (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 15 - 19 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 20 - 24 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 25 - 29 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 30 - 34 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 35 - 39 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 40 - 44 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> 45 - 49 Years (Per Thousand Females):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> Chinese (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> Malays (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Total Fertility Rate -> Indians (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Gross Reproduction Rate (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Net Reproduction Rate (Per Female):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Crude Birth Rate (Per Thousand Residents):
Data prior to 1980 pertain to total population. Data from 1980 onwards pertain to resident population (i.e. Singapore citizens and permanent residents).
Resident Live-Births (Number):
Refers to births with at least one parent who is a Singapore citizen or permanent resident. Data for 1980-1985 are estimated.

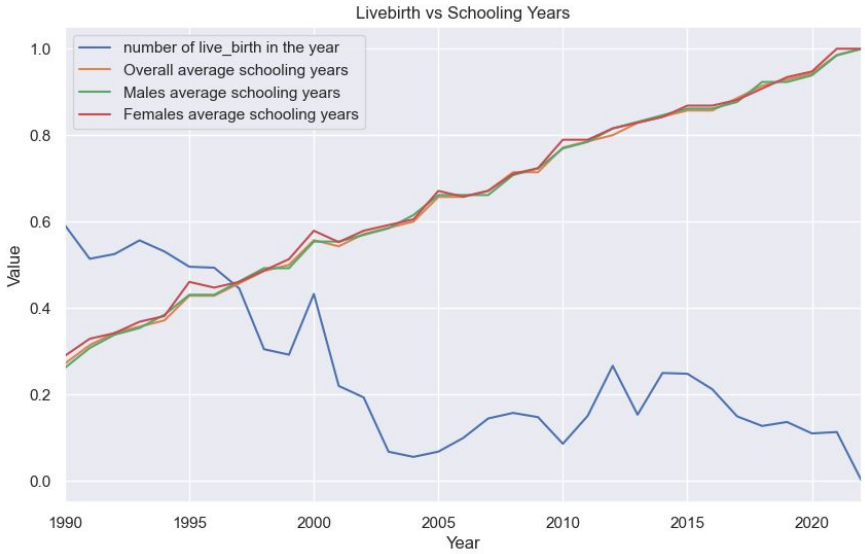
The Datasets



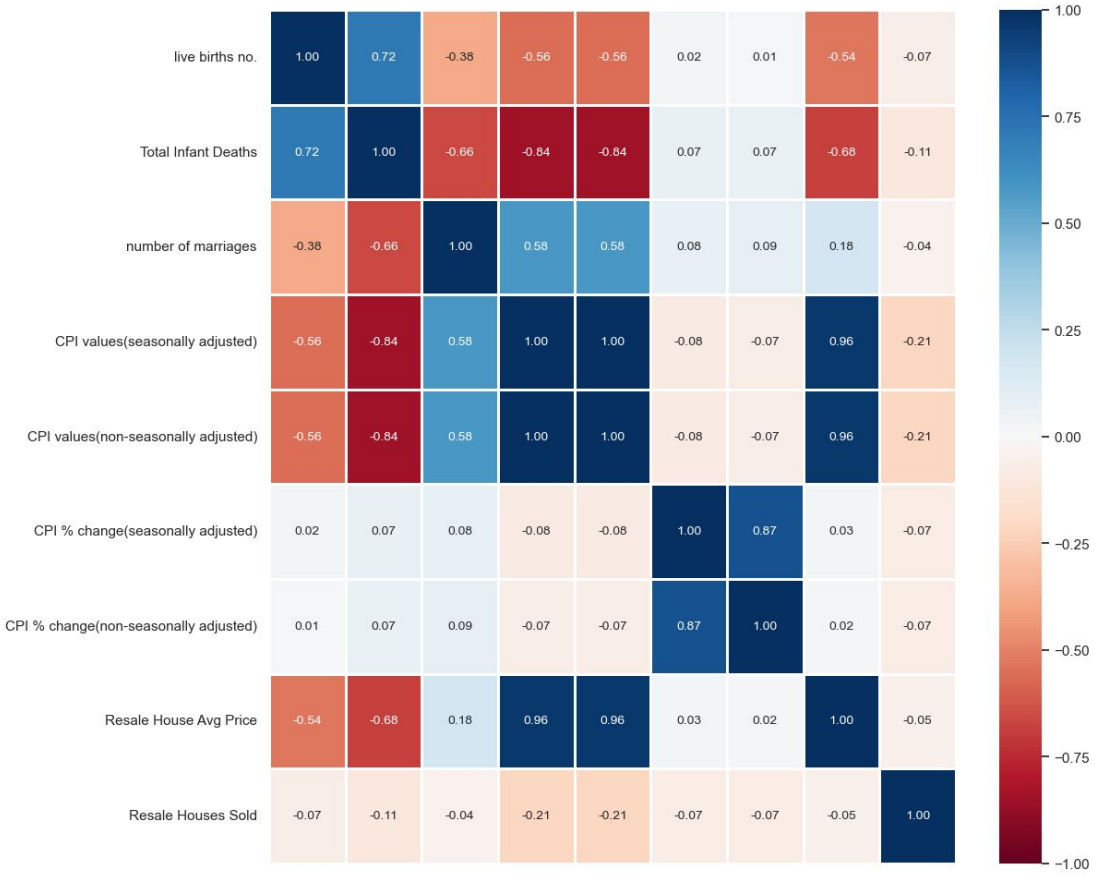
03
Exploratory Data Analysis

Visualising the Data





Visualising the Data

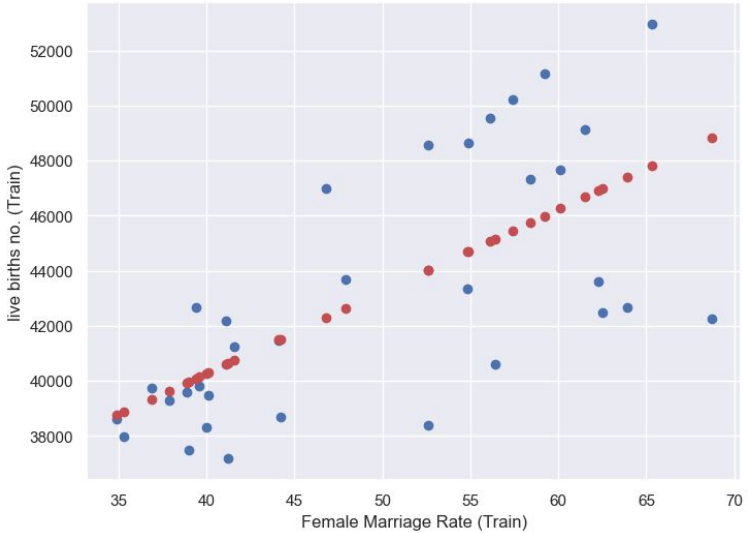


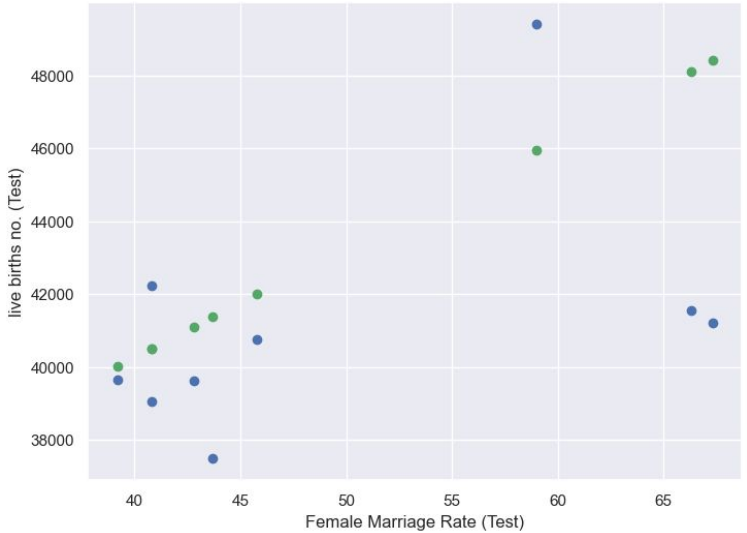
Visualising the Data

	Correlation
Fertility Rate	0.85
Total Infant Deaths	0.77
Male Marriage Rate	0.72
Total Residents	-0.66
Unemployment Rate	-0.59
Crude Marriage Rate	0.56
CPI Values	-0.55
Avg Resale House Price	-0.54

Monthly Dataset Yearly Dataset

Initial Modelling





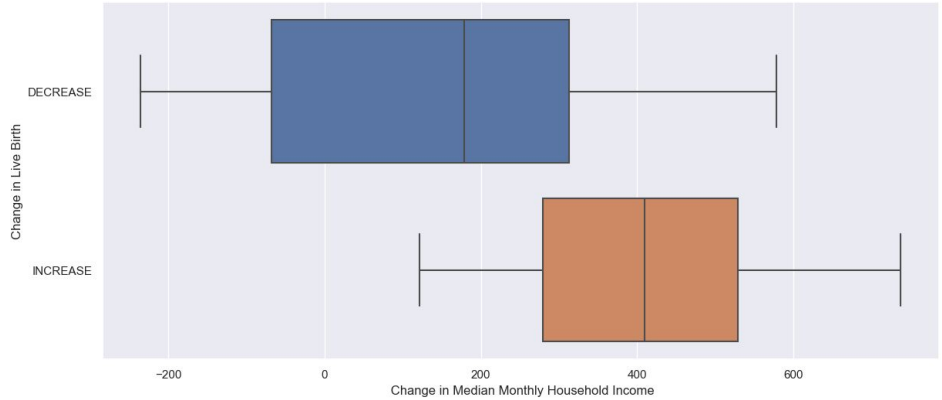
	Train Dataset	Test Dataset
Explained Variance	0.5757	-10.2208
Mean Squared Error	9459	2741

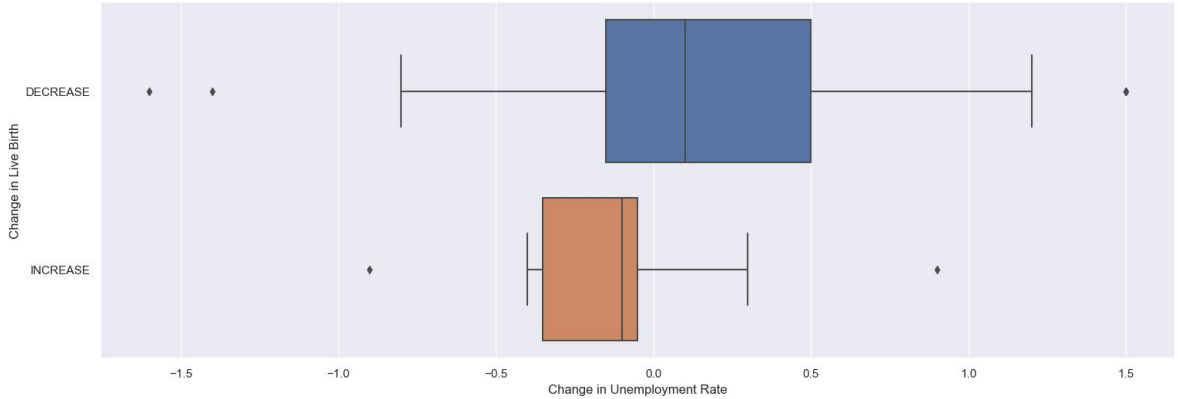
The change to classification problem

	Year	number of live_birth in the year	Total Infant Deaths	number of marriages	CPI values	Inflation(CPI % change from previous year)	Resale House Avg Price	Resale Houses Sold
0	1960-01-01	61775	2158	NaN	NaN	NaN	NaN	NaN
1	1961-01-01	59930	1937	3664.0	24.401	NaN	NaN	NaN
2	1962-01-01	58977	1843	4561.0	24.513	0.5	NaN	NaN
3	1963-01-01	59530	1674	5366.0	25.033	2.1	NaN	NaN
4	1964-01-01	58217	1738	6112.0	25.448	1.7	NaN	NaN

	Year	Change in Live Birth	number of live_birth in the year	Total Infant Deaths	number of marriages	CPI values	Inflation(CPI % change from previous year)	Resale House Avg Price	Resale Houses Sold	Total Population (Number)	...	Female poly & uni 25_39	Poly and Uni Total	Median Income	Median Monthly Household Income
1	1961-01-01	DECREASE	59930	1937	3664.0	24.401	NaN	NaN	NaN	1702400	...	NaN	NaN	NaN	NaN
2	1962-01-01	DECREASE	58977	1843	4561.0	24.513	0.5	NaN	NaN	1750200	...	NaN	NaN	NaN	NaN
3	1963-01-01	INCREASE	59530	1674	5366.0	25.033	2.1	NaN	NaN	1795000	...	NaN	NaN	NaN	NaN
4	1964-01-01	DECREASE	58217	1738	6112.0	25.448	1.7	NaN	NaN	1841600	...	NaN	NaN	NaN	NaN

The change to classification problem





04 Machine Learning Analysis

Machine Learning Models

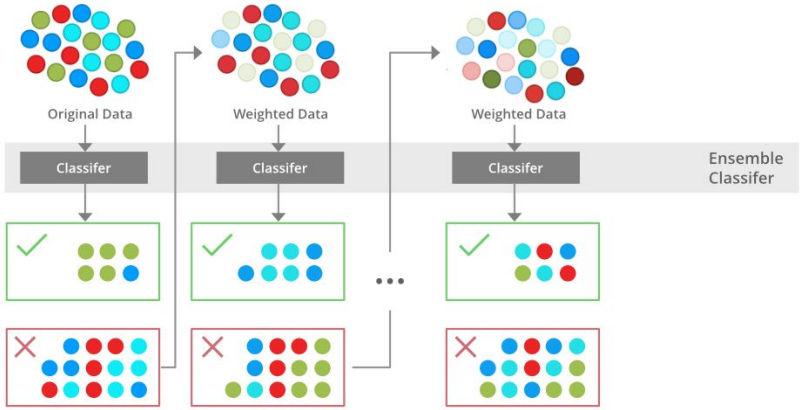
XGBoost

Random Forest

K-Nearest Neighbors

XGBoost

- Iterative procedure built using ensemble decision tree
- Combines the predictions of multiple weak models to produce a stronger prediction
- Efficient handling of missing values



Improving the Dataset

Raw Data
57.36%

Change Data
87.28%

% Change Data
87.82%

	Year	Change in Live Birth	number of live_birth in the year	Total Infant Deaths	number of marriages	CPI values
1	1961-01-01	DECREASE	59930	1937	3664.0	24.401
2	1962-01-01	DECREASE	58977	1843	4561.0	24.513
3	1963-01-01	INCREASE	59530	1674	5366.0	25.033
4	1964-01-01	DECREASE	58217	1738	6112.0	25.448

	Year	Change in Live Birth	Change in Live Birth Number	Change in Total Infant Death	Change in number of marriages	Change in CPI	Change in Inflation
1	1961-01-01	DECREASE	-0.029866	-0.102410	NaN	NaN	NaN
2	1962-01-01	DECREASE	-0.015902	-0.048529	0.244814	0.004590	NaN
3	1963-01-01	INCREASE	0.009377	-0.091698	0.176496	0.021213	3.200000
4	1964-01-01	DECREASE	-0.022056	0.038232	0.139023	0.016578	-0.190476

XGBoost Hyperparameter Tuning

Monthly Dataset

Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = 6 gamma = 0 eta = 0.3	max_depth = 5 gamma = 0.05 eta = 0.15	max_depth = 6 gamma = 0.1 eta = 0.15
	61.64%	63.44%	68.16%

Yearly Dataset

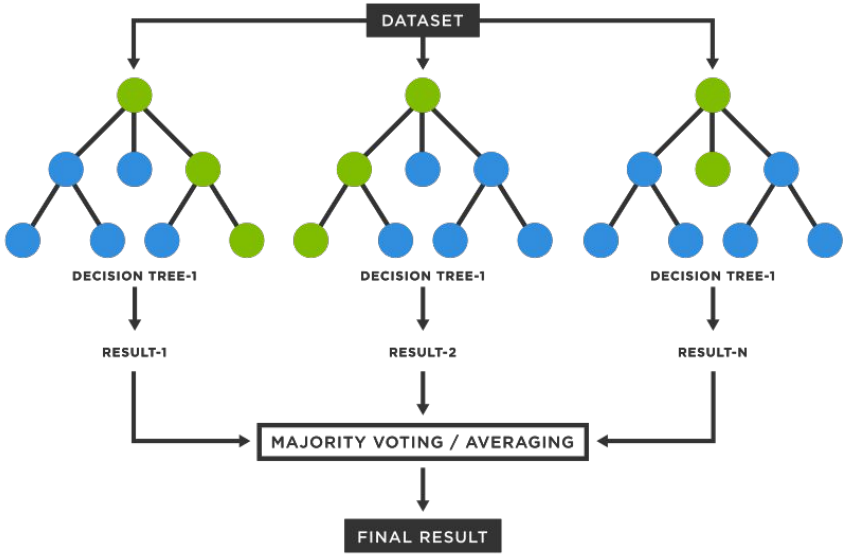
Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = 6 gamma = 0 eta = 0.3	max_depth = 3 gamma = 0.2 eta = 0.25	max_depth = 3 gamma = 0.3 eta = 0.25
	88.48%	89.12%	92.85%

Combined Dataset

Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = 6 gamma = 0 eta = 0.3	max_depth = 3 gamma = 0.2 eta = 0.25	max_depth = 3 gamma = 0.1 eta = 0.1
	87.82%	90.23%	95.55%

Random Forest

- Ensemble learning method for classification
- Combines multiple decision trees to create a more accurate and robust model
- Reduces overfitting by averaging multiple decision trees.



Random Forest Hyperparameter Tuning

Monthly Dataset

Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = None max_features = sqrt n_estimators = 100	max_depth = 5 max_features = 5 n_estimators = 1000	max_depth = 9 max_features = 2 n_estimators = 200
	62%	63.87%	66%

Yearly Dataset

Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = None max_features = sqrt n_estimators = 100	max_depth = 3 max_features = 14 n_estimators = 1000	max_depth = 6 max_features = 27 n_estimators = 50
	73.68%	84.21%	90.79%

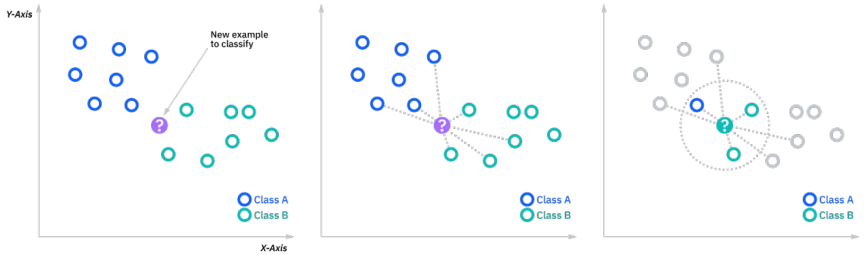
Combined Dataset

Accuracy	Default parameters	Selected parameters	GridSearchCV
	max_depth = None max_features = sqrt n_estimators = 100	max_depth = 3 max_features = 17 n_estimators = 1000	max_depth = 3 max_features = 33 n_estimators = 70
	68.42%	78.94%	93.01%

K-Nearest Neighbors

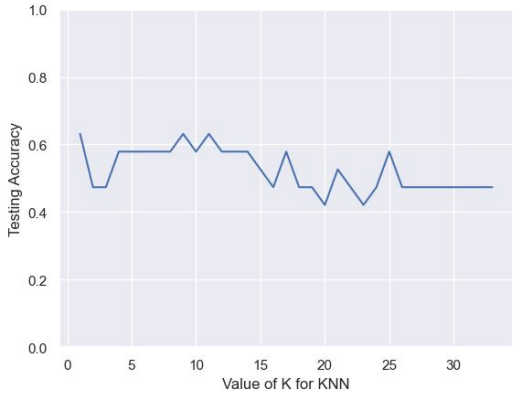
- Non parametric, supervised learning classifier

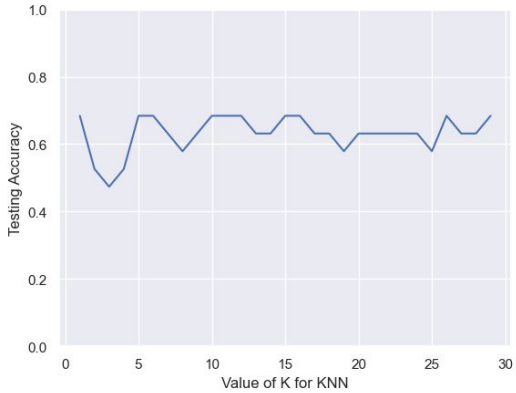
- Uses proximity to make classifications or predictions



K-NN Feature Reduction

		K-NN (Default)	K-NN (After Feature Reduction)
Accuracy (Best)	Monthly Dataset	57%	66%
	Yearly Dataset	58%	73%
	Combined Dataset	69%	79%





05
Results and Findings

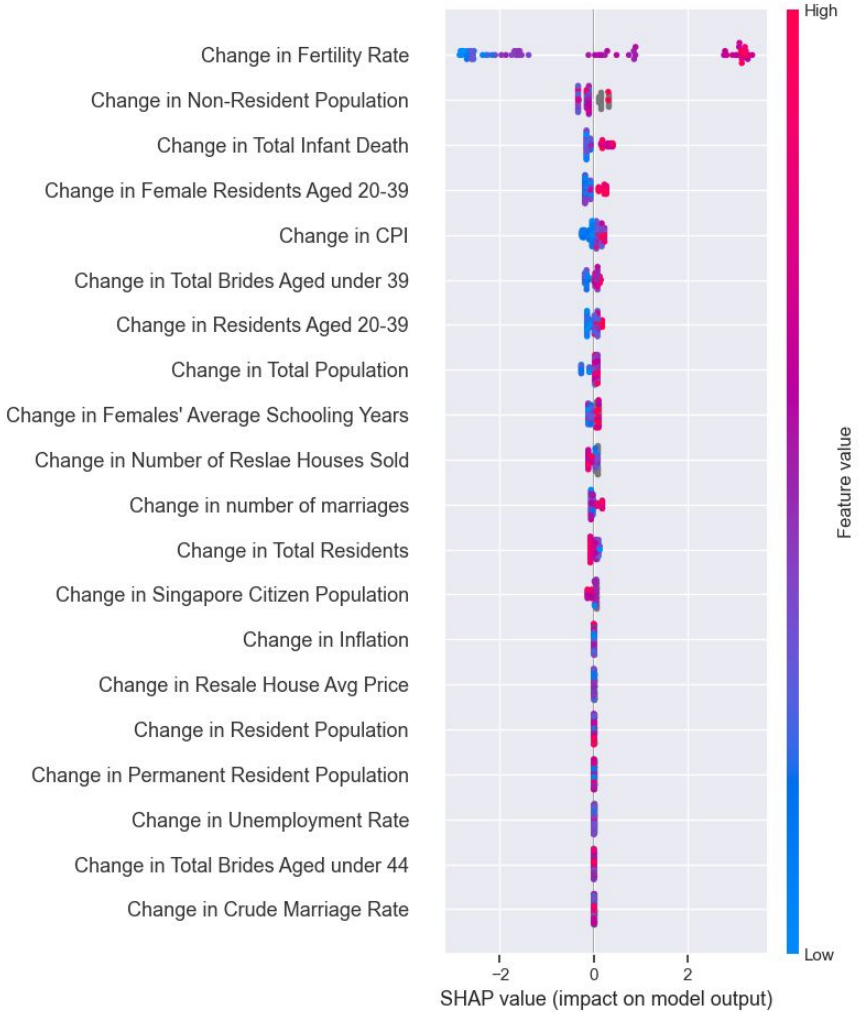
Model Comparison

		XGBoost (GridSearchCV)	Random Forest (GridSearchCV)	K-Nearest Neighbors (After Feature Reduction)
Accuracy	Monthly Dataset	68.16%	66%	66%
	Yearly Dataset	92.85%	90.79%	73%
	Combined Dataset	95.5%	93.01%	79%

Feature Evaluation

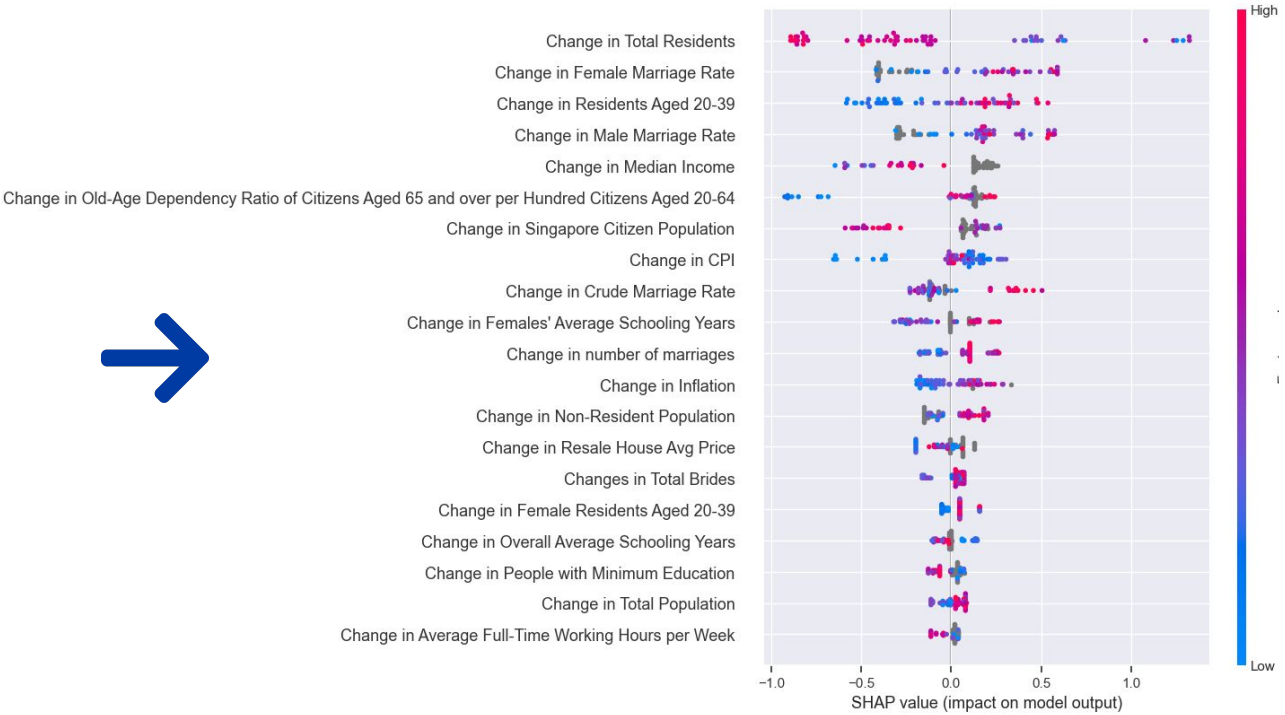
- SHAP (SHapeley Additive exPlanations)

- LIME (Local Interpretable Model-agnostic Explanations




Change in Fertility Rate
Change in Non-Resident Population
Change in Total Infant Death
Change in Female Residents Aged 20-39
Change in CPI
Change in Total Brides Aged under 39
Change in Residents Aged 20-39
Change in Total Population
Change in Females' Average Schooling Years
Change in Number of Reslae Houses Sold
Change in number of marriages
Change in Total Residents
Change in Singapore Citizen Population
Change in Inflation
Change in Resale House Avg Price
Change in Resident Population
Change in Permanent Resident Population
Change in Unemployment Rate
Change in Total Brides Aged under 44
Change in Crude Marriage Rate
High
Low
−2
0
2
SHAP value (impact on model output)

95.55%


Change in Total Residents
Change in Female Marriage Rate
Change in Residents Aged 20-39
Change in Male Marriage Rate
Change in Median Income
Change in Old-Age Dependency Ratio of Citizens Aged 65 and over per Hundred Citizens Aged 20-64
Change in Singapore Citizen Population
Change in CPI
Change in Crude Marriage Rate
Change in Females' Average Schooling Years
Change in number of marriages
Change in Inflation
Change in Non-Resident Population
Change in Resale House Avg Price
Changes in Total Brides
Change in Female Residents Aged 20-39
Change in Overall Average Schooling Years
Change in People with Minimum Education
Change in Total Population
Change in Average Full-Time Working Hours per Week
High
Low
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)

78.95%

Feature Selection

Monthly Dataset	Yearly Dataset	Combined Dataset
Number of Marriages	Total Residents	Residents Aged 20-39
Number of Resale Houses Sold	Female Marriage Rate	Female Marriage Rate
Inflation Rate	People with Minimum Education	Crude Marriage Rate
Average Resale House Price	Median Income	Median Income

Recommendations

- Inflation Rate
- Median Income
- Residents Aged 20-39
- Female Marriage Rate
- Number of Resale Houses Sold

Limitations

- Lack of Data points

- Various missing values (NaN)