GE6226 GIS Research Project

Spatio-Temporal Patterns and Factors Affecting HDB Resale Prices in Singapore

Academic Project Submitted in Fulfilment for The Master of Science

(Applied Geographic Information Systems)

Department of Geography

Faculty of Arts and Social Sciences

National University of Singapore

Lim Zhu An

July 2022

Table of Contents

List of Tables by Pages

List of Figures by Pages

Executive Summary

There are many factors that may affect the resale price of a Housing Development Board (HDB) flat. It is essential to not only understand the spatial patterns in resale prices, but to account for temporal patterns as well. Using a machine learning and a space-time cube data mining approach, this study aims to identify the top factors that affect the resale prices of HDB flats and analyze the spatio-temporal patterns of average resale prices in Singapore. The HDB dataset used for this study was obtained from the Singapore government data portal and consist of all transaction data of HDB resale flats between 1990 to 2022.

The HDB dataset was first geocoded into a point feature class to be used for the training of a supervised random-forest regression model. Distances between each HDB point and nearby spatial features was calculated and used as explanatory variables in the model training. After a suitable model is trained, the top important variables are identified and discussed. In addition, all data points were aggregated into a space-time cube layer by averaging resale prices into a 100m-by-100m fishnet grid at bins of yearly intervals. The space-time cube data was then used to visualize hot or cold spots of resale prices across Singapore for the entire time period.

The results from the regression model shows that the most important variable that affects HDB resale prices is the year of transaction. Similar to previous studies, this study also identified that housing features such as the flat type, floor area, and the age of property are among the top few variables that affect HDB resale prices. The distance variables to nearby features (e.g., Taxi stops, Parks, etc.) have the least importance in predicting HDB resale prices. Overall, there is a statistically significant increasing trend of average resale prices over time. Most HDB estate towns are identified as some type of hot spot category, with the most common type being an oscillating hot spot.

1. Introduction

1.1. Project Context

HDB flats are a common type of property often bought and sold in the Singapore property market. Understanding the spatio-temporal patterns and factors affecting the HDB resale prices will be helpful for potential buyers and sellers of HBD resale flats in the Singapore property market. The use of Geographic Information Science (GIS) and machine learning models have been used in previous studies such as Cao, K. et al. (2019) and Wang, T. et al. (2021) to predict HDB resale prices. These studies have identified several important variables that affects the prediction of HDB resale prices such as the age of property, floor area, date of transaction, and type of flat. Besides the spatial and property features that are analyzed in these types of study, the influence of temporal patterns in resale prices is often overlooked. It is important to look into the temporal trends in parallel to the spatial trends, since HDB resale prices are known to fluctuate over both time and space.

Using different GIS tools, this study aims to analyze the spatio-temporal patterns and factors affecting the changes in HDB resale prices in Singapore. More than 30 years of HDB resale transaction data will be used to develop a supervised random-forest regression model as well as a space-time cube model. The random-forest model will be used to predict HDB resale prices and also determine the top important variables that contributed to the model predictions. The space-time cube model will be used to identify the different types of hot or cold spots of average resale prices from 1990 to 2022 across the study area.

1.2. Term of Reference

1.2.1. Project Scope

The first objective of this study is to determine the top factors that affect HDB resale prices by using a machine learning model approach. This study will train a supervised random-forest regression model[1] to predict HDB resale prices based on a number of different explanatory variables. After a reliable model is developed, the top important variables that contributed to the model prediction will be identified. This study will then discuss the top important factors identified and their relationship with the predicted resale prices. The model will be validated using a test dataset where the Root Mean Square Error (RMSE) and R^2 will be evaluated to determine the accuracy of the model and ensure that the trained model is not overfitted.

The second objective of this study is to analyze the spatial-temporal patterns of average resale prices from 1990 to 2022 across the study area. The Arcgis space-time cube data mining tool[2] will first be used to aggregate the different transactions of resale prices over time within a specific spatial and temporal neighborhood. All HDB resale data points will be aggregated into space-time cube bins which will then be used for analysis to identify hot or cold spots of average resale prices across Singapore. The results of the hot spot analysis will be presented as a map and the different statistical trends (increasing or decreasing) in average resale prices will be assessed for every bin.

[1] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/forestbasedclassificationregression.htm
[2] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm#:~:text=Creating%20a%20space%2Dtime%20cube,2D%20and%203D%20visualization%20techniques

1.2.2. Study Area

The study area for this project is Singapore, and it encompasses all transactions of HDB flats that were built and sold between 1990 to 2022. This study only analyzes the prices of resale HDB flats and does not include private properties such as condominiums and bungalows. Not all towns in Singapore have HDB flats, and the locations of all the transacted HDB resale flats within the study area are shown in Figure 1 below.

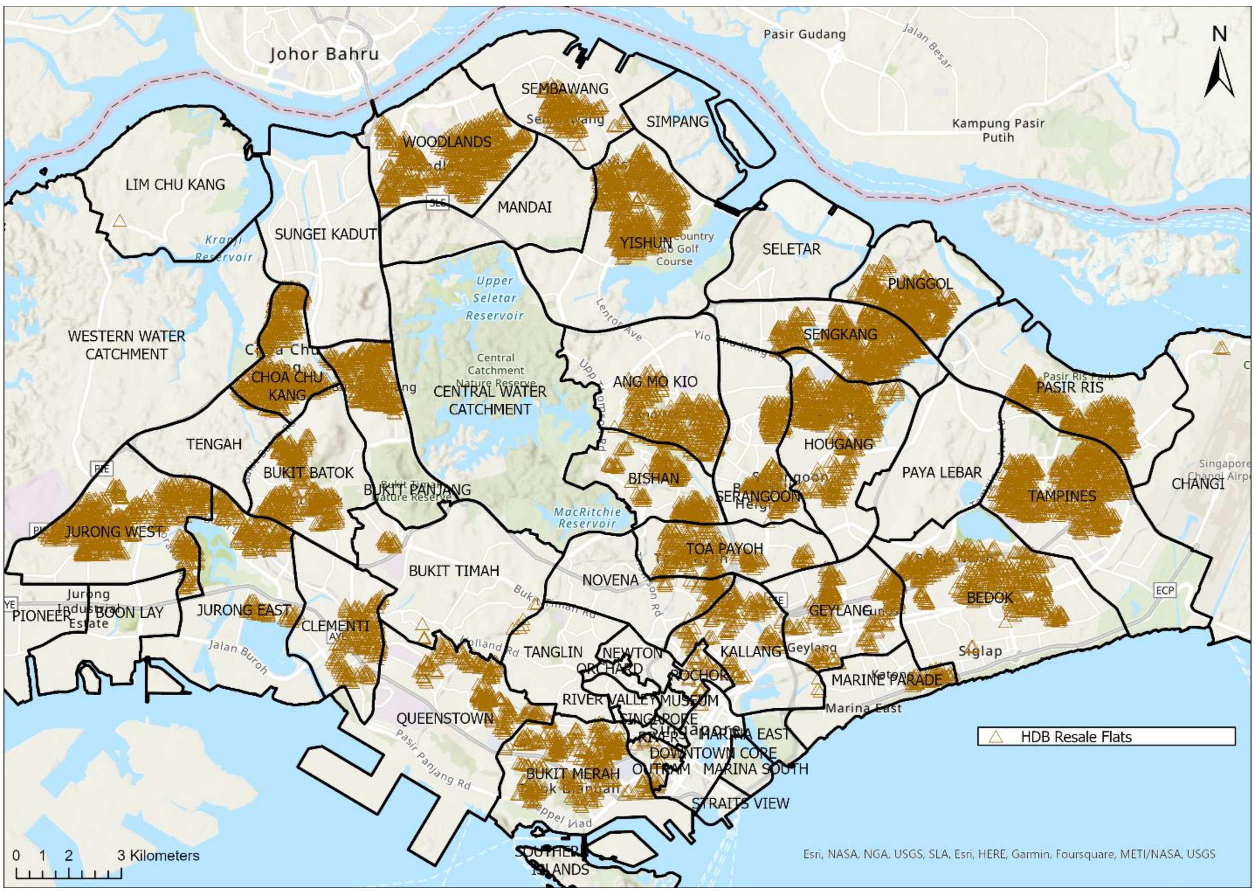

Figure 1: Locations of all HDB resale flats within Study area

2. Data and Methodology

2.1. Data cleaning and pre-processing

All data required for this study are obtained from Singapore government data portals[3] and are summarized in Table 1 below. The primary dataset that is used for this study is the HDB resale prices dataset retrieved from data.gov.sg which includes all transactions of HDB resale flats between 1990 and 2022. The resale prices dataset was first merged into a single dataframe to be processed and contains 866,566 entries with information such as the resale price, flat type, lease commence date, etc. (refer to Table 2). The block and street name columns were first combined to get the address of each resale flat, which is then geocoded using python script and Onemap API[4] to obtain the latitude and longitude for each data point. After geocoding and data cleaning to ensure that the resale flat coordinates correspond with the town label, 854,549 data points remain. These point features will be the main dataset used for the different analysis in the following sections.

Table 1: Summary of all spatial data layers

No.	Data Layer	Data Format	Type
1	Taxi Stops	KML	Point
2	Mass Rapid Transit (MRT) Stations	SHP	Point
3	Parks	KML	Polygon
4	Park Connectors	SHP	Polyline
5	Hawker Centers	KML	Point
6	Shopping malls/Hotels/Places of Interest (POIs)	CSV	Point
7	Schools	SHP	Point
8	Pre-Schools	KML	Point
9	Nature Reserve	SHP	Polygon
10	Water Bodies	SHP	Polygon
11	Water Activities	SHP	Point
12	Bus Stops	SHP	Point

[3] www.data.gov.sg & www.datamall.lta.gov.sg
[4] https://www.onemap.gov.sg/docs/

A new column “property age” was created by subtracting the current year with the lease commence date, and the remaining lease column was dropped due to large amount of missing data. The column for storey range contains 25 categories and will affect the performance of the model. Therefore, the data was regrouped into 5 categories based on their different storey range; 1 to 10, 10 to 20, 20 to 30, 30 to 40, 40 and above. The column for flat model was dropped as well for the same reason.

Table 2: Columns in HDB resale prices dataset

No.	Data Column	Data Type	Remarks
1	month	String	Converted to Datetime object
2	town	String	
3	flat_type	String	E.g., one-room, two-room, etc.
4	block	String	
5	street_name	String	
6	storey_range	String	25 unique categories
7	floor_area_sqm	Int	
8	flat_model	String	20 unique categories
9	lease_commence_date	Int	
10	remaining_lease	String	Dropped due to 709,050 missing values
11	resale_price	Int	In Singapore Dollars (SGD)

Similar to the HDB resale dataset, the data for shopping malls, hotels, and POIs are geocoded using the same methodology. A total of 215 rows were successfully geocoded from the original 314 rows, and were converted into a point feature class. All spatial layers were then exported to a Arcgis geodatabase as feature classes to be used for further analysis, and were all projected using the SVY21 projected coordinate system to reduce errors for distance calculations.

2.2. Project Methodology and Workflow

In order to examine the importance of different variables in predicting HDB resale flat prices, feature engineering had to be done to create explanatory spatial variables from the spatial layer inputs by calculating the nearest distance between each transacted flat and each spatial feature using the Arcgis proximity tool[5] (Figure 2). This study only considers Euclidean distances between features and does not account for a travel time-based spatial matrix which was discussed in the study by Cao, K. et al. (2019). In total, there are 8 explanatory training variables (i.e., flat_type, floor_area_sqm, property_age, storey, year, month, X, Y) and 12 explanatory training distance variables (based on the different spatial layers in Table 1) used for the model training. The dataset will be first split into 70% for training and 30% for test validation. The test data set will not be included in the model training. After the model training is completed and the validation test is satisfactory, the top important variables will be ranked and identified.

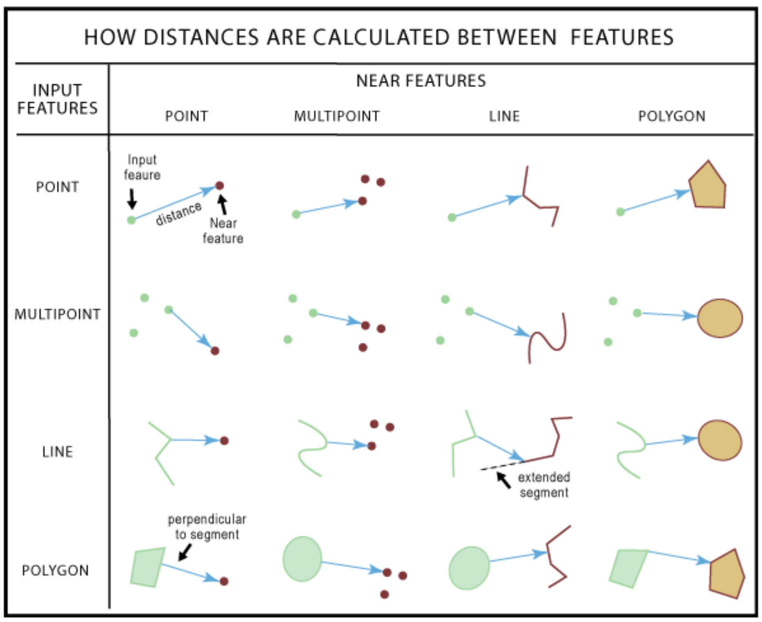

Figure 2: Proximity tool for calculating distance between spatial features

After the random-forest model results is analyzed, the HDB point data will be used to generate a netCDF space-time cube layer for further spatio-temporal analysis. The space-time cube data will consist of aggregated data points in 100m-by-100m fishnet grids across the study area, and the resale prices of each transaction within each grid is aggregated into time bins at a

[5] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/analysis/how-near-analysis-works.htm

yearly interval to get the average resale price (Figure 3). 100m^2 is the estimated size of a HDB block, hence all data points within a block distance will be aggregated into fishnet grids. Using the space-time cube, an emerging hot spot analysis[6] will be done to identify different types of hot or cold spots of average resale prices within a neighborhood distance of 500m across the study area. This neighborhood distance was selected as it covers several blocks of HDB flats and will show if there are any significant hot or cold spots of resale prices within that neighborhood area.

The hot spot analysis tool will calculate the Getis-Ord Gi* statistics[7] for each feature in the dataset and will indicate whether there is a significant hot or cold spot at any given location based on its neighbors. For the temporal trend analysis in the space-time cube data, the tool relies the Mann-Kendall statistics[8] to evaluate if the time steps show significant increasing, decreasing, or no trend (Hamed, K. H., 2009). The statistics of the space-time cube and the results of the hot spot analysis will be discussed in the follow sections.

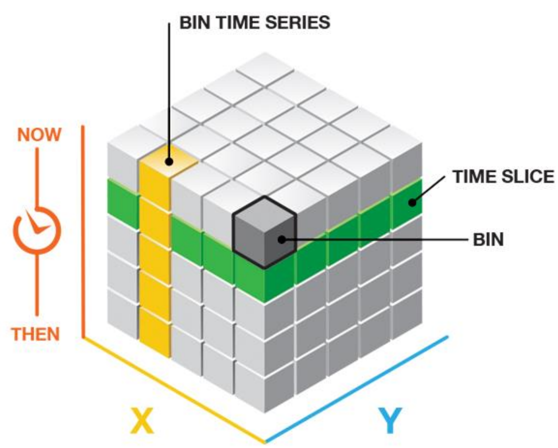

Figure 3: Space-time Cube Illustration

[6] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/emerginghotspots.htm
[7] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm
[8] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm

3. Results

3.1. Discussion

3.1.1. Supervised Random-forest Model

The model characteristics and model results are summarized in Appendix 1. The top important feature that affects HDB resale prices is the year of transaction of the flat, and it contributes to approximately 42% of all explanatory variables. This suggests that there is a strong temporal influence over the trend of resale prices. As seen in Figure 4 below, there is a clear increasing trend in the resale prices over time. This increasing trend in resale prices will be further assessed using the space-time cube data and emerging hot spot analysis in the following section. The increase in HDB resale prices over time is expected due to inflation and limited land space in Singapore for housing estates.

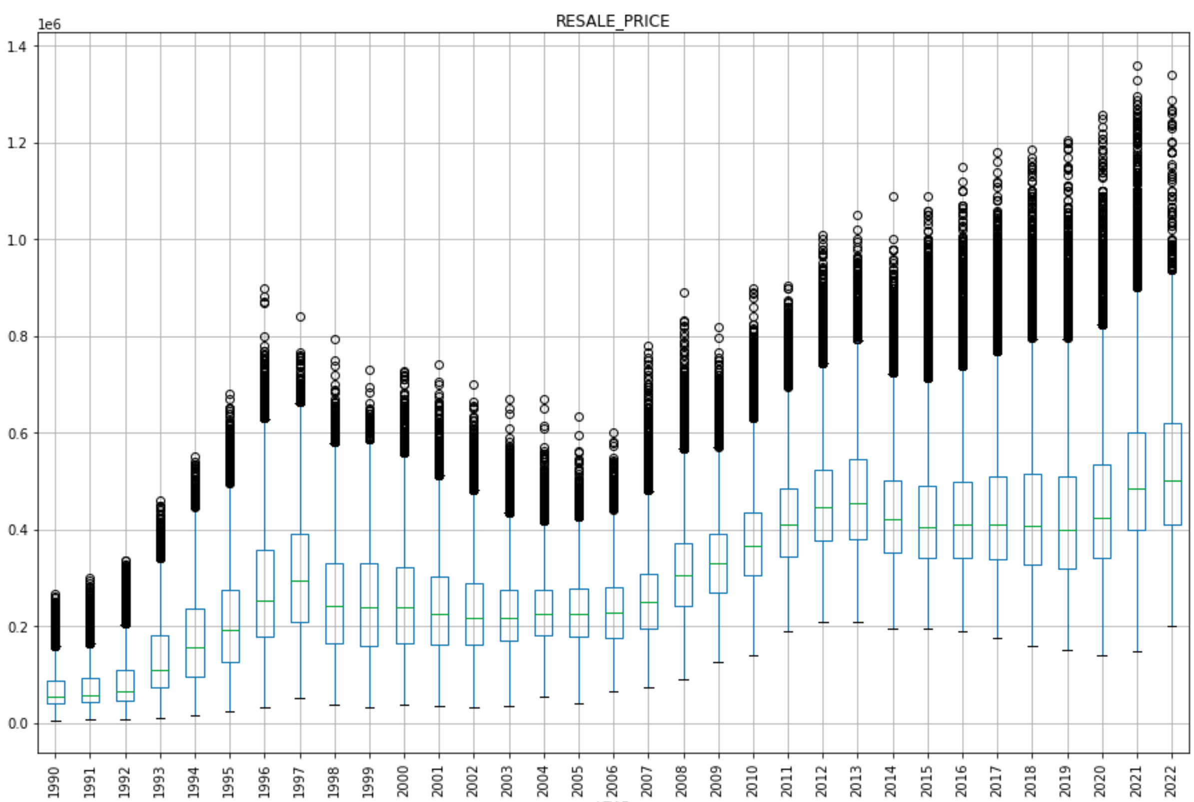

Figure 4: Boxplot showing yearly statistics of resale prices from 1990 to 2022

The second, third, fourth, and seventh important variables are; floor area, flat type, property age, and storey. These variables account for 20%, 15%, 8%, and 2% respectively of all explanatory variable importance. These variables identified in this study are similar to the hedonic pricing models discussed in the studies by Cao, K. et al. (2019) and Wang, T. et al. (2021). These property features are often considered by buyers and sellers in the HDB resale property market, which directly influence the resale prices. Spatial location of the HDB resale flat affects the resale price as well. The Latitude and Longitude variables are the fifth (4%) and sixth (2%) important variables respectively. Although they only contribute to 6% of the variable importance in total, they are still ranked in the top 10 important variables of the model.

Distances from HDB flat to nearby features are least important as compared to the other types of variables mentioned above. Among all the variables, there are 8 distance variables identified with 1% importance respectively, and they are ranked in the following order: 8) Nature reserve, 9) Schools, 10) Taxi stops, 11) Watersport facilities, 12) Hawker centers, 13) MRT stations, 14) Park connectors, 16) POIs. The location of these spatial features relative to the HDB resale flats is shown in Figure 5 below. Distance to parks, water bodies, pre-schools, and bus stops have less than 1% importance. Unexpectedly, the distance to bus stops is the least important distance variable. This was not expected as bus is a common form of transport that is often used by residents in HDB estates.

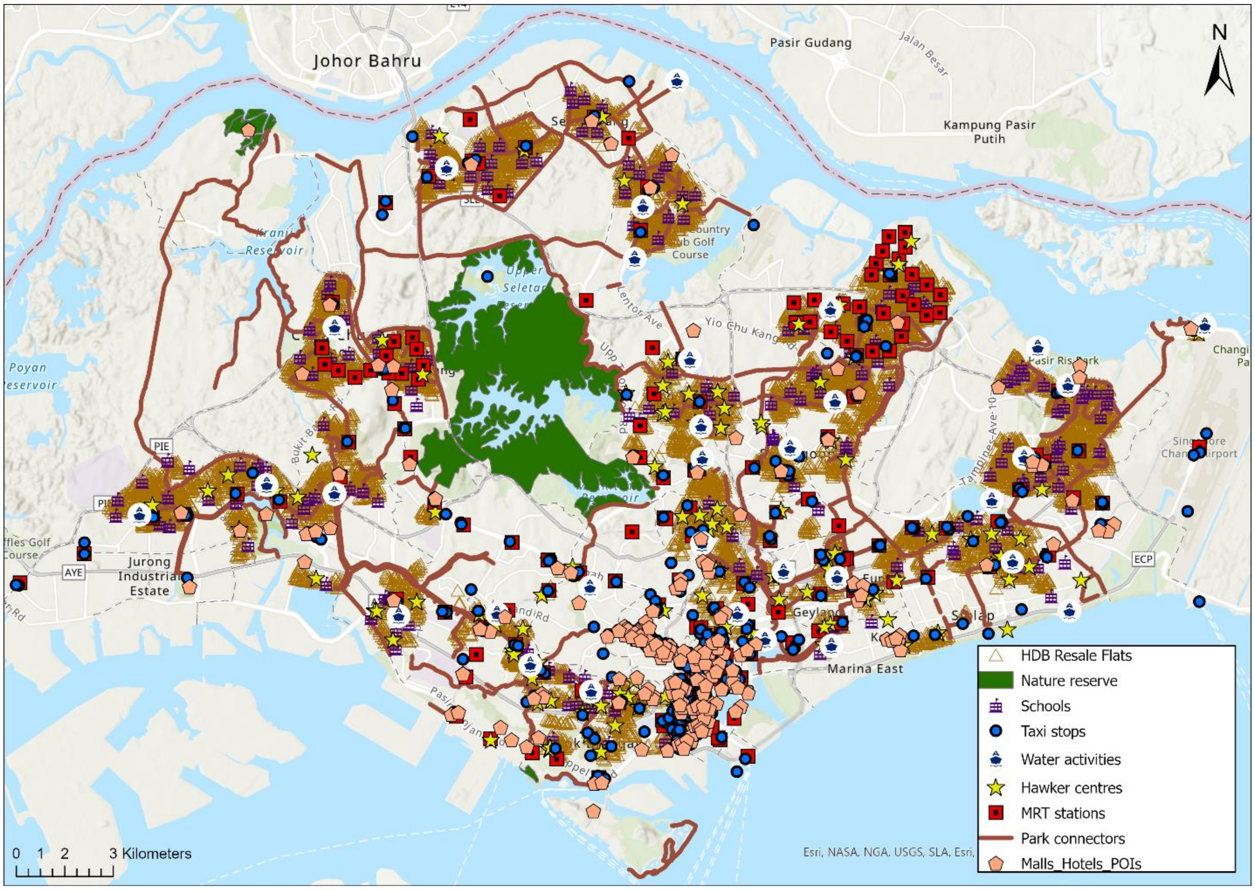

Figure 5: Top important spatial features

3.1.2. Emerging Hot spot Analysis

To further analyze the spatio-temporal trend of the resale prices, a hot spot analysis was conducted to identify different types of hot and cold spots of average resale prices based on a yearly interval from 1990 to 2022. The results of the emerging hot spot analysis are published as an interactive web map[9]. Figure 6 shows the summary of all the hot and cold spots across Singapore, and Figure 7 shows the overall statistical trend result for each space-time cube bin. The different categories of hot and cold spots are listed in Table 4 (Appendix 4). As expected, most HDB estate towns are identified as some type of hot spot such as new, consecutive, sporadic, oscillating, historical, or intensifying hot spot. The results for the areas that do not have

[9] https://www.arcgis.com/apps/mapviewer/index.html?webmap=88bb4a6a4c0244d29a2684624fd7b866

any HDB resale flats sold for the entire time period are not accessed, and those bins are assigned with zero value.

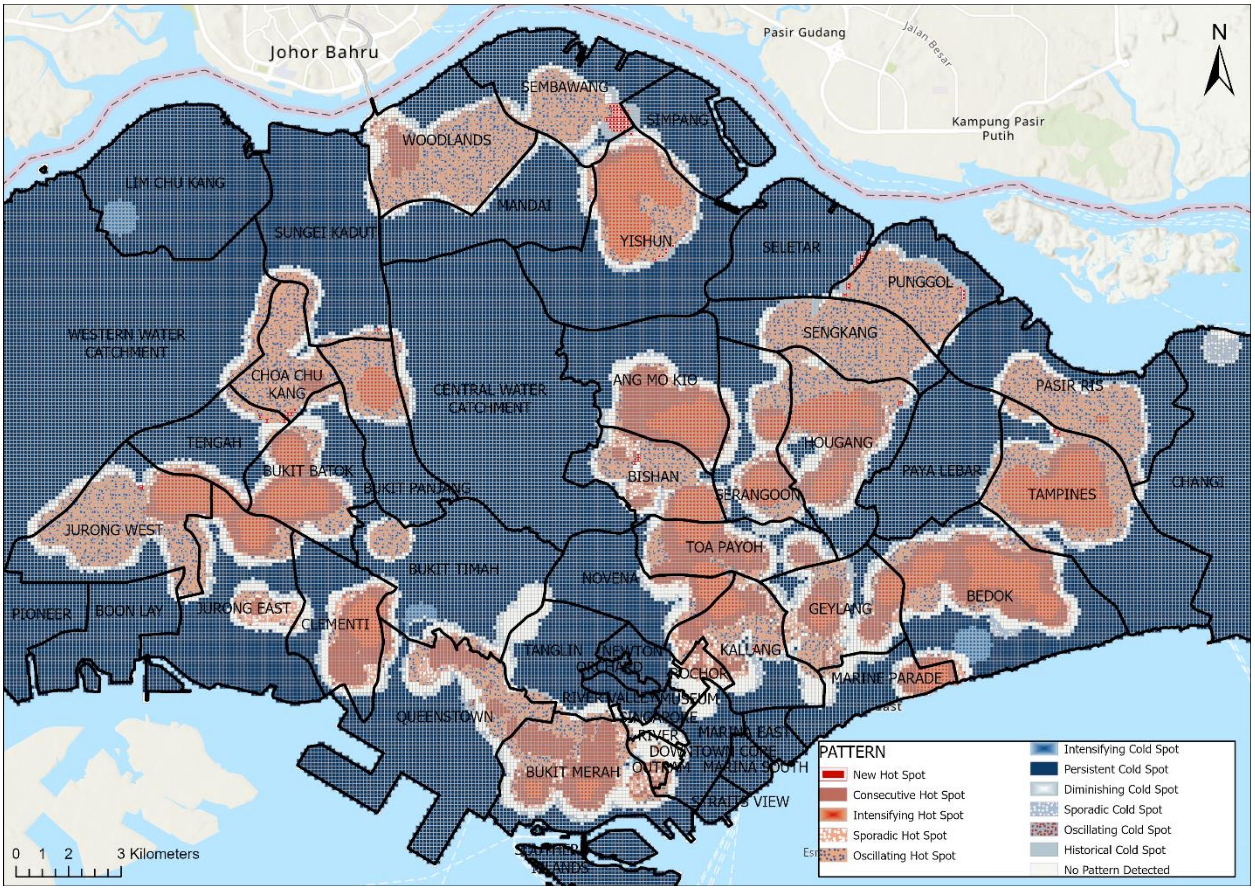

Figure 6: Emerging Hot spot Analysis Results

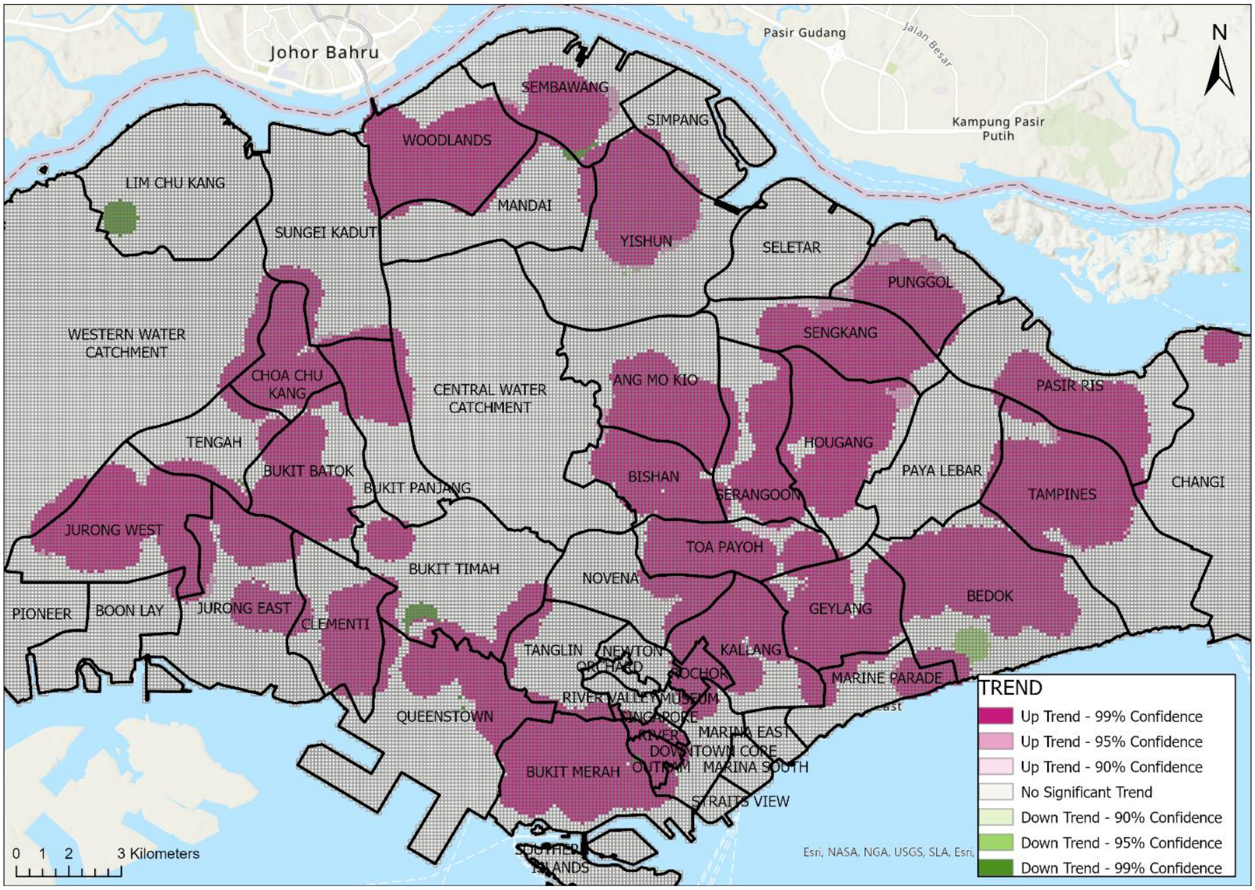

Figure 7: Statistical Trend Map

The most common type of hot spot identified is an oscillating hot spot, which accounts for approximately 51% of all types of hot spot categories. Majority of the HDB estate towns across Singapore were identified with oscillating hot spots. This means that they are a statistically significant hot spot for the last time-step but has a history of also being a statistically significant cold spot during a prior time-step. Sembawang, Punggol, Sengkang, and Pasir Ris, have the largest percentage of oscillating hot spot in their respective town zones as compared to other types of hot spot categories.

The location identified with the largest area of new hot spot is Canberra estate located in Sembawang. The average resale prices in that area are hot for the first time in the most recent time step interval and has never been a statistical hot spot before. This is most likely because the estate is relatively new and was built in 2017, hence resale of those HDB flats only begin recently (i.e., 2021 to 2022).

Most of the HDB estate towns have intensifying hot spots with the exception of Punggol, Sengkang, Woodlands, Sembawang, Choa Chu Kang, and Queenstown. This indicates that most HDB estate towns have at least 90% of the time step intervals as statistically significant hot, and are becoming hotter over time. In addition, the intensity of clustering of high average resale prices in each time step is increasing overall and that increase is statistically significant. Besides this, most towns are also identified with having consecutive hot spots.

Apart from hot spots, there are some locations with cold spots as well. There are three areas with intensifying cold spot identified and they are located at Lim Chu Kang, Bukit Timah/Queenstown, and Bedok. This indicates that the average resale prices in these locations have been a statistically significant cold spot for 90% of the time-step intervals, including the final time step. They are also calculated to have a downward trend with 90% to 99% confidence as shown in Figure 7 below. The location at northern Changi is also identified as a sporadic cold spot.

3.2. Evaluation

To evaluate the performance and accuracy of the random-forest model, 10 iterations of validation on the test data set is performed and the results are summarized in Table 3 below. The median R^2 for the validation of the model is 0.98 and is statistically significant when comparing the predicted values to the observed values for the test data. This indicates that the trained model performs well even when testing against unknown data in the test data set. The out of bag errors shows that the random-forest model is able to explain about 97.62% of the variation in the model with 100 trees, and indicates that it is a robust model for predicting HDB resale prices based on the set of explanatory variables.

Table 3: Validation data results

OID	R^2	SEED
1	0.979643	121958
2	0.979727	671155
3	0.979839	131932
4	0.979753	365838
5	0.97964	259178
6	0.979735	644167
7	0.979542	110268
8	0.980574	732180
9	0.980408	54886
10	0.980263	137337

Figure 8 shows the scatter plot of the random-forest model predicted resale prices against actual resale prices in the test data set, with a R^2 of 0.98. The random-forest model is capable of predicting resale prices on test data with a RMSE of 22138.88. The validation result of the random-forest model on the test data set is very good considering both the RMSE and R^2 of the test data validation. However, a disadvantage of a tree-based model is that it tends to have poor generalization for cases outside the range of the train set (Wang, T. et al., 2021). For example, the model might have issues predicting resale prices for year 2023 onwards since the upper bound of the train set is only 2022.

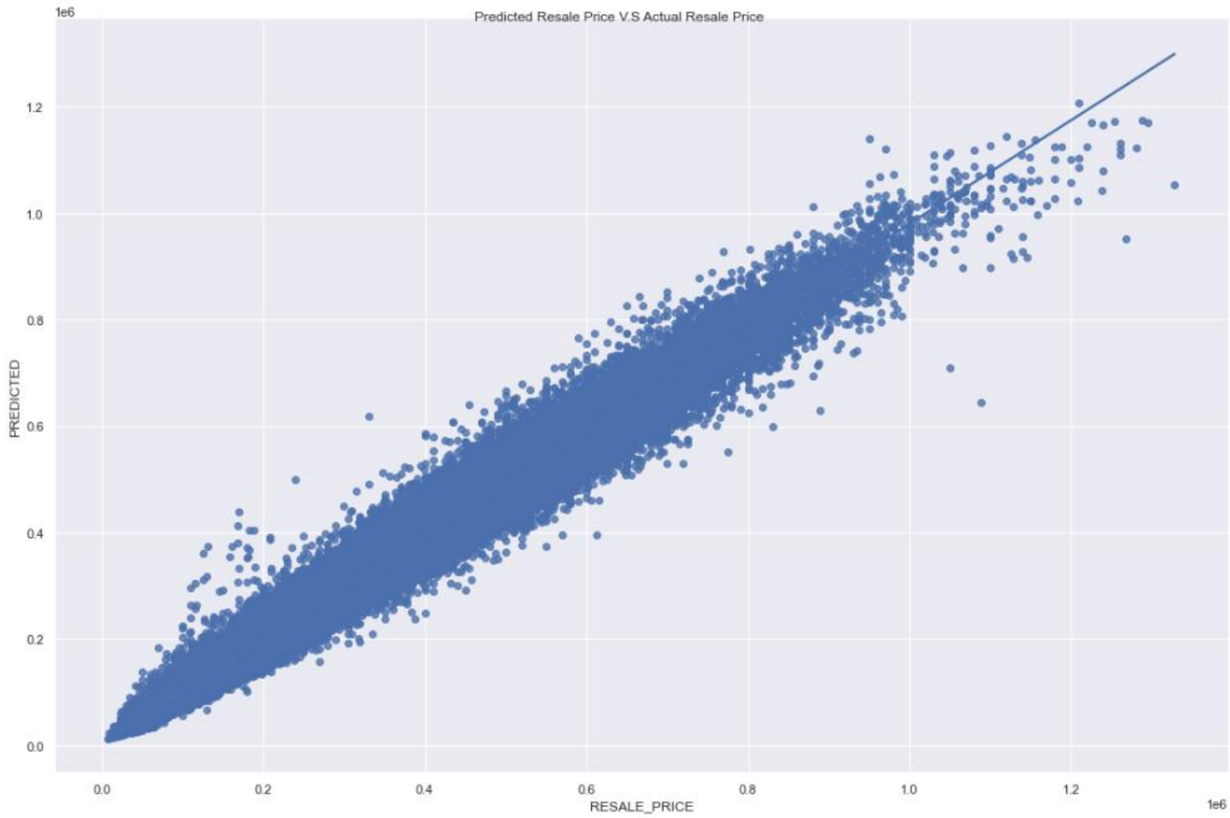

Figure 8: Model predicted resale prices compared to actual resale prices

The space-time cube characteristics and statistics are summarized in Appendix 2. The results shows that there is a statistically significant increase in average resale prices across the entire time period. The overall z-score value for the Mann-Kendall statistics is 6.96 with a p-value of less than 0.05 for all the space-time cubes. The details and statistics of the emerging hot spot analysis are summarized in Appendix 3. The total number of locations analyzed for the hot spot analysis is 68457 grids (around 6.85km^2) and 33 time-steps (i.e., 33 years). Out of all those locations, 15976 (around 1.60km^2) are identified as a type of significant hot spot category.

4. Conclusion

The year that which the HDB resale flat was transacted is the most important factor that affects resale prices. Temporal influence on resale property value should be the primary consideration for buyers or sellers when they plan to buy or sell their HDB property. Based on the space-time cube and hot spot analysis, the resale prices of HDB flat shows a statistically significant increasing trend from 1990 to 2022. Similar to previous studies done, this study also found that housing property features such as floor area, flat type, property age, and storey, are important features that affect the resale price. The distance from HDB flats to nearby features are far less important as compared to the temporal and property variables.

Considering that this study did not account for spatially varying relationship between the data, future studies might consider using a geographically weighted regression model with the given set of spatial features and HDB resale data. In addition, instead of using only Euclidean distances between spatial objects, a travel time distance approach might yield different results. The limitation to the hot spot analysis and space-time cube approach is that it does not take into account that the transaction of HDB resale flats do not occur every year. The type and number of units that are sold in each block also affects the average resale price in the space-time cube calculation as well.

Total Number of words: 4106

References

1. Cao, K., Diao, M., & Wu, B. (2019). A big data-based geographically weighted regression model for public housing prices: A case study in Singapore. Annals of the American Association of Geographers, 109(1), 173-186. https://doi.org/10.1080/24694452.2018.1470925
2. Hamed, K. H. (2009). Exact distribution of the Mann–Kendall trend test statistic for persistent data. Journal of Hydrology (Amsterdam), 365(1), 86-94. https://doi.org/10.1016/j.jhydrol.2008.11.024
3. Wang, T., Wang, Y., & Liu, M. (2021). A price prediction method based on CatBoost. Paper presented at the 403-408. https://doi.org/10.1109/ICCST53801.2021.00090

Appendices

Appendix 1

Random Seed: 42

Model Characteristics

Number of Trees	100
Leaf Size	5
Tree Depth Range	44-55
Mean Tree Depth	48
% of Training Available per Tree	100
Number of Randomly Sampled Variables	6
% of Training Data Excluded for Validation	30

Model Out of Bag Errors

Number of Trees	50	100
MSE	636167461.477	574290959.447
% of variation explained	97.365	97.621

Top Variable Importance

Variable	Importance	%
year	6002061967368141.00	42
floor_area_sqm	2845385173205931.50	20
flat_type	2133522685444196.25	15
property_age	1108125472425711.25	8
Y	636580643777519.88	4
X	297389120586259.19	2
storey	235783175703415.88	2
G_MP08_NAT_RES_PL	207209971192709.47	1
DUS_SCHOOL_SPORTS_FACILITIES	112207465213478.95	1
TAXISTOPS_PROJECTED	108548158869505.14	1
AQUATICSG	94026626945601.92	1
HAWKERCENTRES_PROJECTED	92319962383452.27	1
MRTLRTSTNPTT	81916076426397.19	1
G_MP08_PK_CONECTR_LI	76907957022382.84	1

month	75373045186613.91	1
POI_PROJECTED	73748618518594.27	1
PARKS_PROJECTED	57816334982292.94	0
G_MP08_WATERBODY_PL	49080381837949.27	0
PRESCHOOLS_PROJECTED	23155732612326.24	0
BUSSTOP	20546805868567.82	0

Training Data: Regression Diagnostics

R-Squared	0.993
p-value	0.000
Standard Error	0.000

*Predictions for the data used to train the model compared to the observed categories for those features

Validation Data: Regression Diagnostics

R-Squared	0.980
p-value	0.000
Standard Error	0.000

*Predictions for the test data (excluded from model training) compared to the observed values for those test features

```
Median R2 0.980 was approximately reached at seed 671155
```

Explanatory Variable Range Diagnostics

Variable	Training		Validation		Share	
	Minimum	Maximum	Minimum	Maximum	Training[a]	Validation[b]
floor_area_sqm	31.00	297.00	31.00	307.00	0.96*	1.04
year	1990.00	2022.00	1990.00	2022.00	1.00	1.00
property_age	3.00	56.00	3.00	56.00	1.00	1.00
X	103.69	103.99	103.69	103.99	1.00	1.00
Y	1.27	1.46	1.27	1.46	1.00	1.00
month	1.00	12.00	1.00	12.00	1.00	1.00
POI_PROJECTED	13.03	3418.31	13.03	3418.31	1.00	1.00
AQUATICSG	42.99	5939.50	42.99	5939.50	1.00	1.00
TAXISTOPS_PROJECTED	21.92	5498.74	21.92	5498.74	1.00	1.00
HAWKERCENTRES_PROJECTED	6.98	7379.40	6.98	7379.40	1.00	1.00

MRTLRTSTNPTT	21.36	5465.63	21.36	5465.63	1.00	1.00
PRESCHOOLS_PROJECTED	0.00	4924.49	0.00	4924.49	1.00	1.00
DUS_SCHOOL_SPORTS_FACILITIES	0.00	5144.60	0.00	5144.60	1.00	1.00
BUSSTOP	8.23	377.43	8.23	377.43	1.00	1.00
G_MP08_PK_CONECTR_LI	3.55	2162.12	3.55	2162.12	1.00	1.00
PARKS_PROJECTED	6.04	2500.89	6.04	2500.89	1.00	1.00
G_MP08_WATERBODY_PL	6.55	1558.77	6.55	1558.77	1.00	1.00
G_MP08_NAT_RES_PL	188.54	17442.79	188.54	17442.79	1.00	1.00

(a) % of overlap between the ranges of the training data and the input explanatory variable

(b) % of overlap between the ranges of the validation data and the training data
* Data ranges do not coincide. Training or validation is occurring with incomplete data.

+ Ranges of the training data and prediction data do not coincide and the tool is attempting to extrapolate.

Appendix 2

The space time cube has aggregated 854549 points into 186148 fishnet grid locations over 33 time step intervals. Each location is a 100 meters by 100 meters square. The entire space time cube spans an area 53800 meters west to east and 34600 meters north to south. Each of the time step intervals is 1 year in duration so the entire time period covered by the space time cube is 33 years. Of the 186148 total locations, 5680 (3.05%) contain at least one point for at least one time step interval. These 5680 locations comprise 187440 space time bins of which 130532 (69.64%) have point counts greater than zero. There is not a statistically significant increase or decrease in point counts over time.

Space Time Cube Characteristics

Input feature time extent	1990-01-01 00:00:00 to 2022-01-01 00:00:00
Number of time steps	33
Time step interval	1 year
Time step alignment	End
First time step temporal bias	100.00%
First time step interval	after 1989-01-01 00:00:00 to on or before 1990-01-01 00:00:00
Last time step temporal bias	0.00%
Last time step interval	after 2021-01-01 00:00:00 to on or before 2022-01-01 00:00:00
Coordinate System	SVY21
Cube extent across space	(coordinates in meters)
Min X	2631.9890
Min Y	15702.5277
Max X	56431.9890
Max Y	50302.5277
Rows	346
Columns	538
Total bins	6142884

COUNT

Total number of locations	**186148**
Locations with at least one point	5680
- associated bins	187440
- % non-zero (sparseness)	69.64

Summary Field - RESALE_PRICE_MEAN_ZEROS

% of locations excluded due to unfilled bins	**0.00%**
- Total number	0
Total number of locations	5680
- associated bins	187440
% of included locations with estimated bins	90.16%
- Total number	5121
% of all bins that were estimated	30.36%
- Total number	56908

Overall Data Trend - COUNT

Trend direction	Not Significant
Trend statistic	-1.1001
Trend p-value	0.2713

Overall Data Trend - RESALE_PRICE_MEAN_ZEROS

Trend direction	Increasing
Trend statistic	6.9570
Trend p-value	0.0000

Appendix 3

Input Space Time Cube Details

Distance interval	100 meters
Time step interval	1 year
Aggregation Shape Type	Fishnet Grid
First time step temporal bias	100.00%
First time step interval	after
	1989-01-01 00:00:00
	to on or before
	1990-01-01 00:00:00
Last time step temporal bias	0.00%
Last time step interval	after
	2021-01-01 00:00:00
	to on or before
	2022-01-01 00:00:00
Number of time steps	33
Number of locations analyzed	68457
Number of space time bins analyzed	2259081
% non-zero	5.78%

Analysis Details

Neighborhood distance	500 meters
Neighborhood time step intervals	1 (spanning 1 year)

Summary of Results

	HOT	COLD
New	84	0
Consecutive	3635	0
Intensifying	3135	238
Persistent	0	46279
Diminishing	0	954

Sporadic	985	684
Oscillating	8137	5
Historical	0	284

All locations with hot or cold spot trends: 64420 of 68457

Category Definitions

Last time step is hot:

- New: the most recent time step interval is hot for the first time
- Consecutive: a single uninterrupted run of hot time step intervals, comprised of less than 90% of all intervals
- Intensifying: at least 90% of the time step intervals are hot, and becoming hotter over time
- Persistent: at least 90% of the time step intervals are hot, with no trend up or down
- Diminishing: at least 90% of the time step intervals are hot, and becoming less hot over time
- Sporadic: some of the time step intervals are hot
- Oscillating: some of the time step intervals are hot, some are cold

Last time step is not hot:

- Historical: at least 90% of the time step intervals are hot, but the most recent time step interval is not

Last time step is cold:

- New: the most recent time step interval is cold for the first time
- Consecutive: a single uninterrupted run of cold time step intervals, comprised of less than 90% of all
- Intensifying: at least 90% of the time step intervals are cold, and becoming colder over time
- Persistent: at least 90% of the time step intervals are cold, with no trend up or down
- Diminishing: at least 90% of the time step intervals are cold, and becoming less cold over time intervals
- Sporadic: some of the time step intervals are cold
- Oscillating: some of the time step intervals are cold, some are hot

Last time step is not cold:

- Historical: at least 90% of the time step intervals are cold, but the most recent time step interval is not

Appendix 4

Table 4: Categories of hot and cold spot[10]

Pattern Type	**Definition**
New	A location that is a statistically significant hot/cold spot for the final time step and has never been a statistically significant hot/cold spot before.
Consecutive	A location with a single uninterrupted run of statistically significant hot/cold spot bins in the final time-step intervals. The location has never been a statistically significant hot/cold spot prior to the final hot/cold spot run and less than ninety percent of all bins are statistically significant hot/cold spots.
Intensifying	A location that has been a statistically significant hot/cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering of high/low counts in each time step is increasing overall and that increase is statistically significant.
Persistent	A location that has been a statistically significant hot/cold spot for ninety percent of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
Diminishing	A location that has been a statistically significant hot/cold spot for ninety percent of the time-step intervals, including the final time step. In addition, the intensity of clustering in each time step is increasing / decreasing overall and that increase / decrease is statistically significant.
Sporadic	A location that is an on-again then off-again hot/cold spot. Less than ninety percent of the time-step intervals have been statistically significant hot/cold spots and none of the time-step intervals have been statistically significant hot/cold spots.
Oscillating	A statistically significant hot/cold spot for the final time-step interval that has a history of also being a statistically significant cold/hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot/cold spots.
Historical	The most recent time period is not hot/cold, but at least ninety percent of the time-step intervals have been statistically significant hot/cold spots.
No Pattern Detected	Does not fall into any of the hot or cold spot patterns above

[10] https://pro.arcgis.com/en/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmoreemerging.htm#GUID-09587AFC-F5EC-4AEB-BE8F-0E0A26AB9230