

ISSS608 Visual Analytics Project Presentation

Group 1
Chia Yong Qing
Choo Mei Xuan
Clara Chua



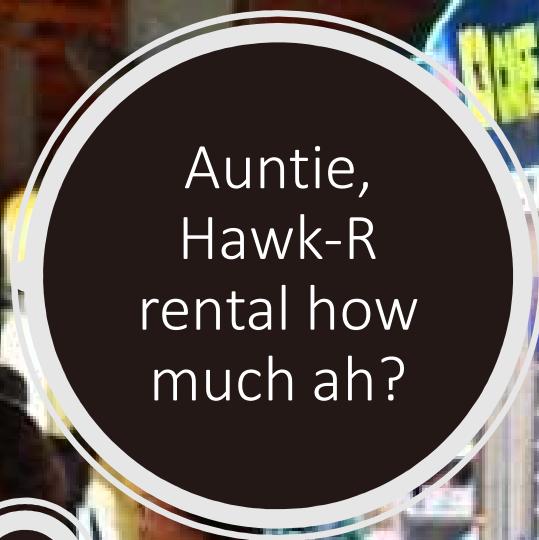
Agenda

- i. Background & Objectives
- ii. Methodology
- iii. Tools and Packages
- iv. App Demonstration
- v. Results and Interpretation

A collage of images depicting a hawker stall and a crowded food center. The top left image shows a stall with red lanterns and a menu board. The bottom left image shows a man in a green shirt serving food. The right side of the collage shows a crowded food center with many people eating at tables.

In July 2018, there was a bid of \$10,028 monthly rental for a drinks stall at Chomp Chomp Food Centre.

Do hawkers know the **market rate** for hawker stalls and **do they know how much to bid?**

A circular callout bubble with a white border and a black background. It contains the following text:

Auntie,
Hawk-R
rental how
much ah?



From This...

TENDERING OF HAWKER STALLS LIST OF SUCCESSFUL TENDERERS FOR SEPTEMBER 2018 TENDER EXERCISE (CLOSED ON 26 SEP 2018)

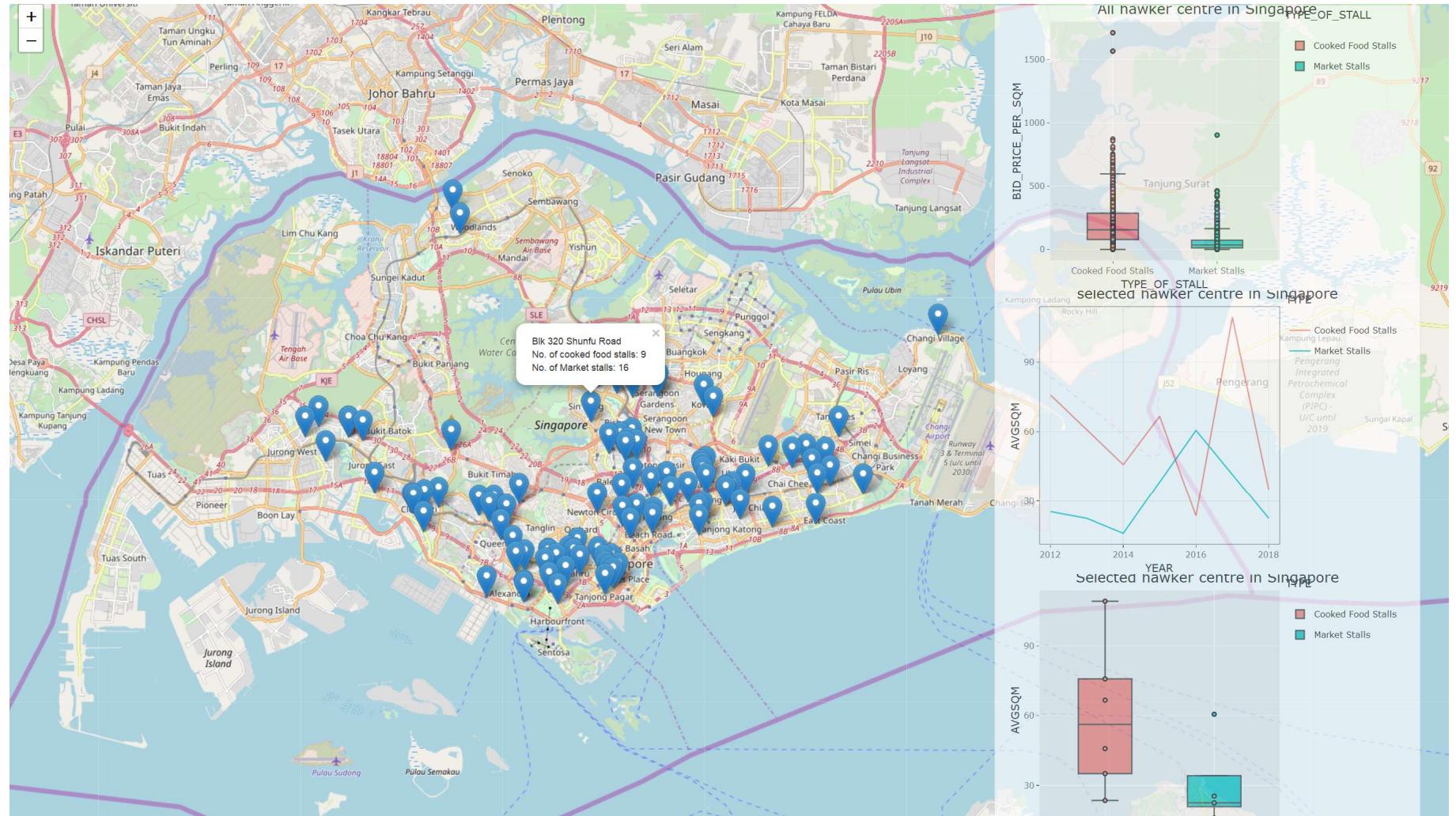
HAWKER CENTRE	STALL NO	ARTICLE OF SALE	TENDERER'S NAME	TENDERED BID
BLK 117 ALJUNIED AVENUE 2	01-66B	PIECE & SUNDRY GOODS	JASININ BINTI MABIN	\$ 257.00
BLK 117 ALJUNIED AVENUE 2	01-151	VEGETABLES	TOH AH SING	\$ 5.00
BLK 226D ANG MO KIO AVENUE 1	01-101	PIECE & SUNDRY GOODS	WONG KIM SENG	\$ 450.00
BLK 341 ANG MO KIO AVENUE 1	01-56	PIECE & SUNDRY GOODS	YEW NIN KONG	\$ 618.00
BLK 628 ANG MO KIO AVENUE 4	01-130	ASSORTED SPICES	FONG PICK SUN	\$ 529.00
BLK 29 BENDEMEER ROAD	01-101	PIECE & SUNDRY GOODS	NOOR HIRNIE BINTE MOHAMED KHIR RAMDAN	\$ 1,018.00
BEO CRESCENT MARKET	01-04	FRESH SEA FOOD	TIM ANDREW YEO ZHI MIN THEODORE	\$ 16.00
BLK 665 BUFFALO ROAD	01-205	PRESERVED & DRIED GOODS	KYAW KYAW HTUN	\$ 463.00
BLK 115 BUKIT MERAH VIEW	01-144	POULTRY	NEO WOON SWEE, NORMEN (LIANG WENRUI)	\$ 400.00
BLK 41A CAMBRIDGE ROAD	01-63	WHOLE FRUITS	TAN CHOON KIAT (CHEN JUNJIE)	\$ 108.00
BLK 41A CAMBRIDGE ROAD	01-64	WHOLE FRUITS	TAN CHOON KIAT (CHEN JUNJIE)	\$ 108.00
BLK 41A CAMBRIDGE ROAD	01-112	FROZEN GOODS	RAHIMA BEEVI BINTE MOHAMUD FARUK	\$ 50.00

NEA website –
“ Notice of
successful
tenders for
September
2018



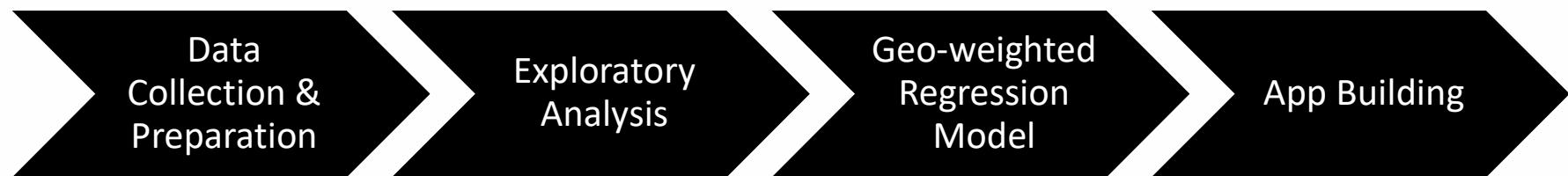
HAWK-R-STALL

To This!



HAWK-R-STALL

Methodology



Column	Description	Example	Data Source
HAWKER_CENTRE	Name of hawker centre	Redhill Hawker Centre	NEA "List Of 5 Highest Tender Bids"
STALL_NO	Unit number of stall	01-66B	NEA "List Of 5 Highest Tender Bids"
STALL_AREA	Size of stall (sqm)	5.65	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
TRADE	Specific type of trade	Vegetables	NEA "List Of 5 Highest Tender Bids"
TRADE_GENERIC	Generic type of trade	Fresh Produce	Derived from "Trade"
TENDERED_BID	Successful bid price (\$)	\$2988	NEA "List Of 5 Highest Tender Bids"
AVERAGE_BID_PRICE	Average successful bid price of Hawker Centre (\$)	\$450	Derived from "Tendered bid"
MONTH	Month of successful bid	Apr	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
YEAR	Year of successful bid	2017	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
DATE	Month and Year of successful bid	Apr-2017	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
TYPE_OF_STALL	Type of business	Market Stalls	NEA website
POSTAL_CODE	6 digits zip code representing postal location of hawker centre	560572	NEA website
LAT	Latitude in 7d.p.	1.2793399	Derived from postal code using R
LONG	Longitude in 7.d.p	103.8466525	Derived from postal code using R
BID_PRICE_PER_SQM	Tendered bid price of stall divided by stall size in sqm	528.85	Derived from tendered price and store size
LAST_RENOVATION	Reopening date after renovation	Feb-2003	NEA Hawker Centres Upgrading Programme
AGE_OF_HAWKER	Number of years past since reopening date	15.8	Derived from fraction of year since reopening date
TYPE_OF_UPGRADING	Reason for upgrading	Reconfiguration	NEA Hawker Centres Upgrading Programme

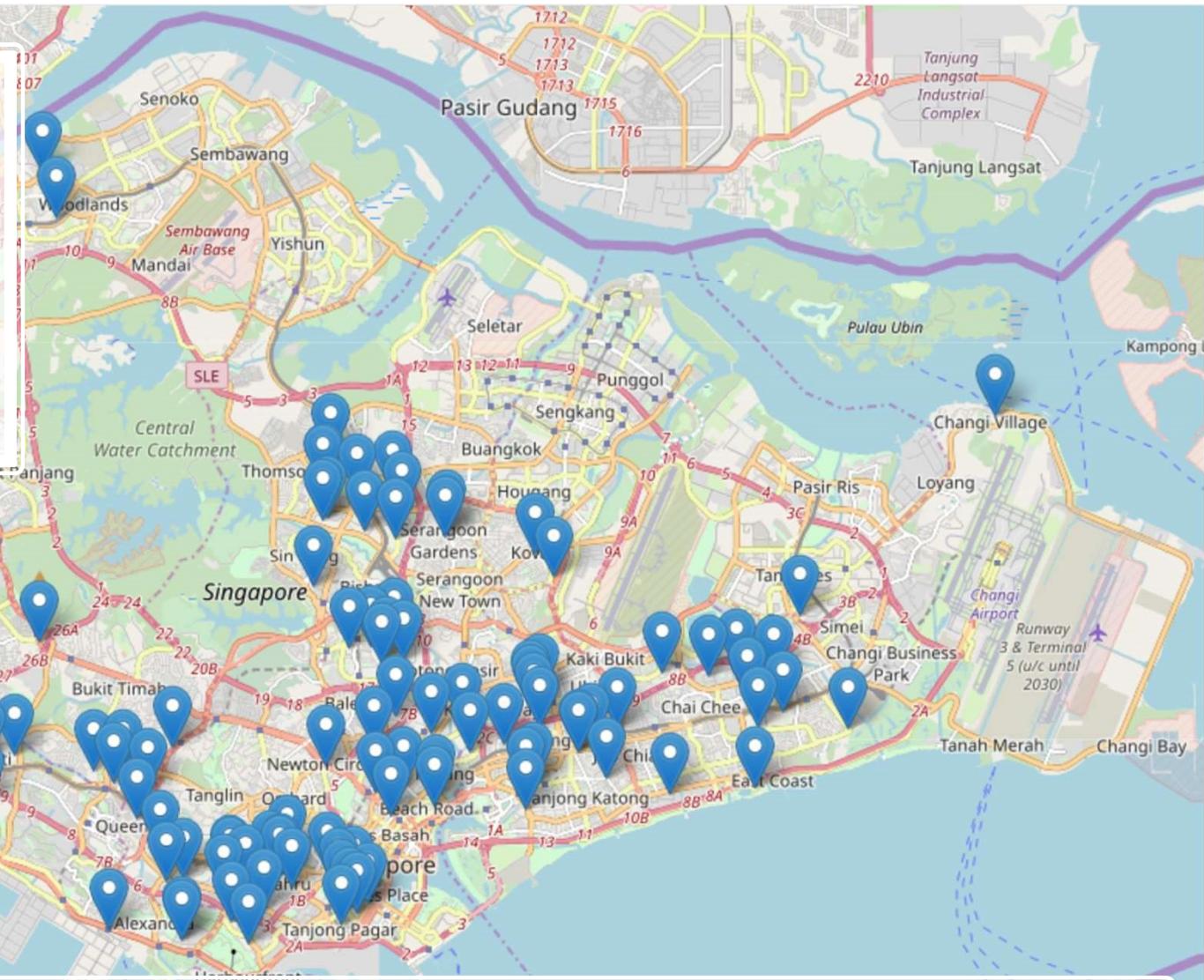
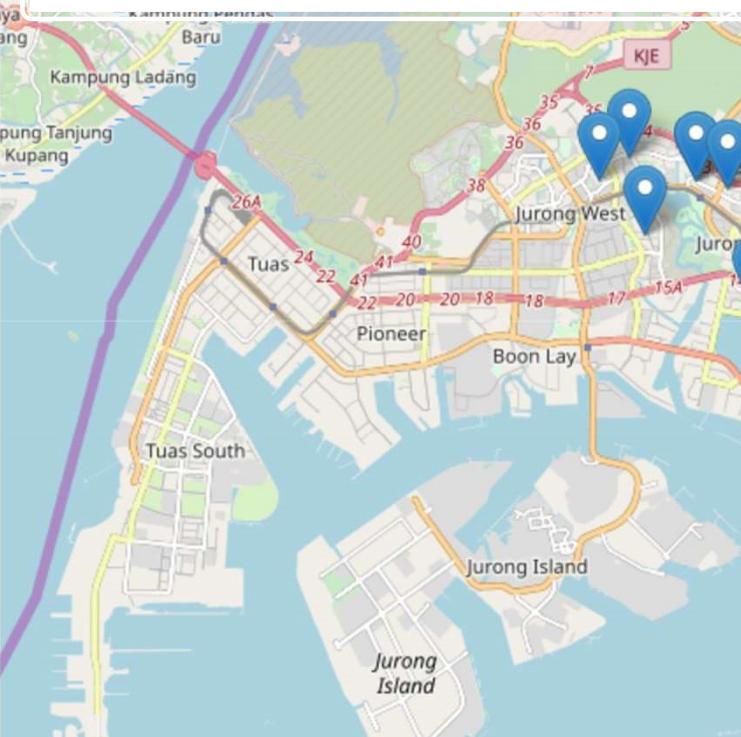
Dataset

- Our project uses data from tender notices published on NEA's website
- Our final working file contains over 4,000 records of successful tender bids from Mar 2012 to Sep 2018.
- Locations information of key variables have been extracted from data.gov.sg and onemap.



HAWK-R-STALL

114 Hawker Centres
Market vs. Food Stalls
>4,000 successful tender
bids **Mar 2012 – Sep 2018**
25 Hawker Trades



Map of NEA regulated hawker centres in Singapore



Tools and Packages



Package	Usage
Shiny	To build interactive web applications with R.
Shinydashboard	Automatically bind reactive inputs and outputs and prebuilt widgets to build responsive applications.
Leaflet	Plots Singapore map from openstreet.sg with panning and zooming functions.
Tidyverse	For usage of ggplot2 visualisations.
Plotly	To add interactivity to ggplot2 charts.
Dygraphs	Allows for plotting of dynamic time series line graph.
Treemap	To plot categorical data using treemap .
Highcharter	To enable interactivity, zoom and export function to treemap.
Sf	To encode spatial vector data.
Sp	For retrieving coordinates and plotting data as map.
GWmodel	Used to calculate optimal bandwidths, and GW regression modelling.
Rgdal	Allows spatial data to be associated with coordinate reference systems.
Tmap	To plot layer-based bubble maps.
Dt	To render results as HTML tables in shiny dashboard.



App Demo

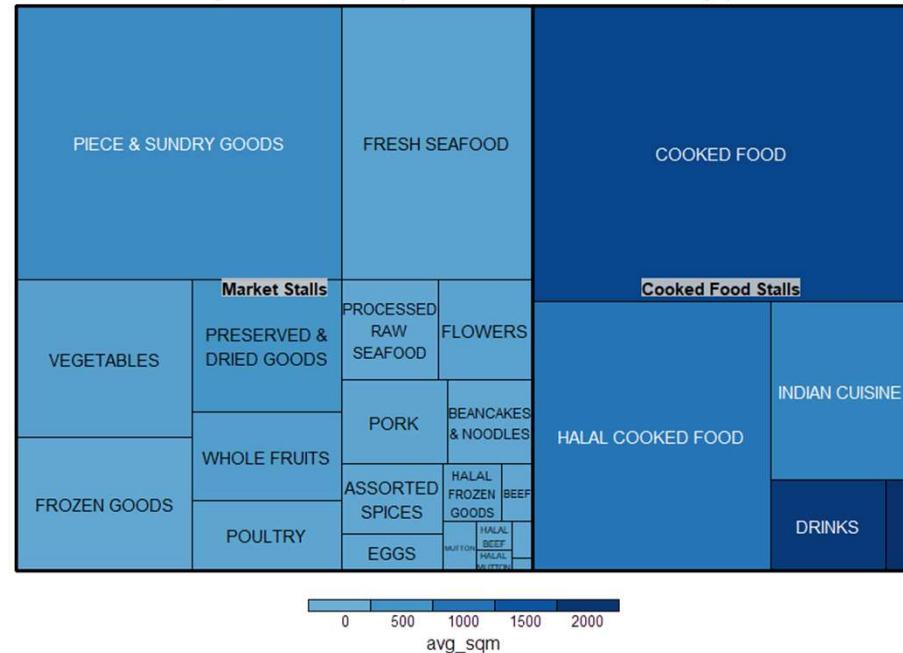


Exploratory Analysis

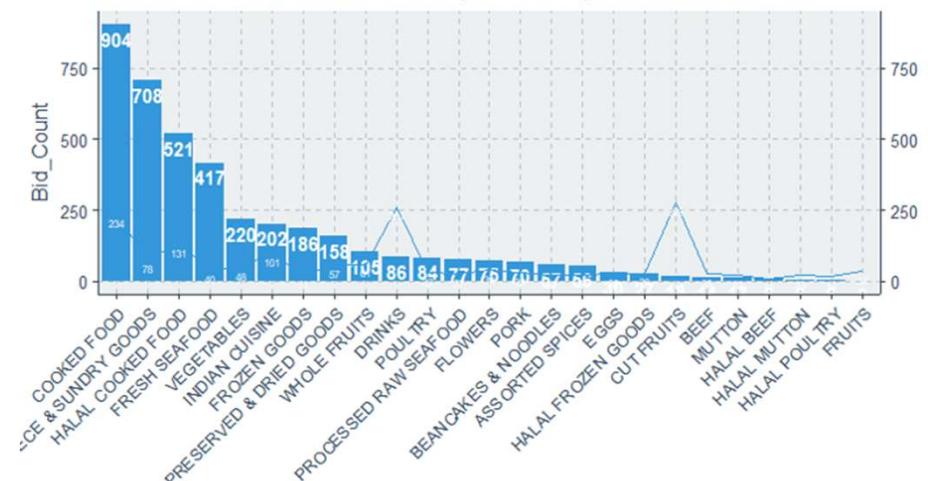
Mar 2012 – Sep 2018

- Collectively, more bid counts for Markets stalls vs. Cooked Food stalls
- Top 3 trades with highest bid count – Cooked food, piece and sundry goods, halal cooked food
- Top 3 trades with highest bid price per sqm: cooked food, drinks, cut fruits

Average Bid Price Per Sqm Across Hawker trades in Singapore



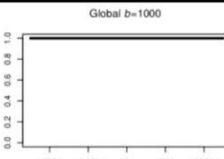
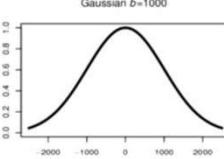
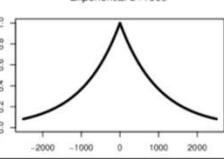
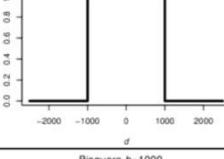
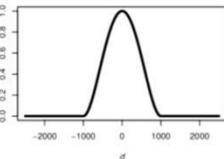
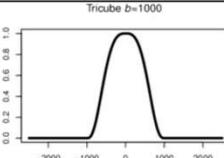
Number of Successful Bid Counts and Average Price Per Sqm of Hawker Trades in SG



Geo-weighted Regression Model

- Useful for dataset where spatial data is not well represented by global model i.e. **local difference is significant**.
- Provides a local model of the rental prices by fitting a regression equation to **every hawker centre** in the dataset.
- Employs **moving window techniques** and **distance decay function** - nearer observations have more weightage in estimating the local set of regression coefficients than observations further away
- It constructs these separate equations by incorporating the dependent and explanatory variables of features falling within the bandwidth of each target feature.

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i$$

Kernel	Formula	Plots	Description
Basic	$w_{ij} = 1$		- Equal weightage for all observations.
Gaussian	$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$		- Continuous function. - Weight decreases according to Gaussian Curve as distance between observation and calibration points increases.
Exponential	$w_{ij} = \exp\left(-\frac{ d_{ij} }{b}\right)$		- Continuous function. - Weight decreases according to Exponential Curve as distance between observation and calibration points increases.
Box Car	$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} < b, \\ 0 & \text{otherwise} \end{cases}$		- Discontinuous function. - Exclude observations that is some distance further from calibration point. - Useful for large dataset.
Bi-Square	$w_{ij} = \begin{cases} (1 - (d_{ij}/b)^2)^2 & \text{if } d_{ij} < b, \\ 0 & \text{otherwise} \end{cases}$		- Discontinuous function. - Weight decreases as distance between observation points increases.
Tri-Cube	$w_{ij} = \begin{cases} (1 - (d_{ij} /b)^3)^3 & \text{if } d_{ij} < b, \\ 0 & \text{otherwise} \end{cases}$		- Discontinuous function. - Weight decreases as distance between observation points increases.

Our dashboard allows user to select from 5 different kernel functions – **Gaussian**, **Exponential**, **Box Car**, **Bi-square**, **Tri-Cube**

With reference to table:
 w_{ij} is the j-th element of the diagonal of the matrix of geographical weights $W(u_i, v_i)$, and d_{ij} is the distance between observations i and j, and b is the bandwidth.



HAWK-R-STALL

Selection of best GWR model

Type of stall	Cooked food			Market		
	Bid price per SQM	Average bid price per SQM	Median bid price per SQM	Model	Global	Basic
AIC	1,142	1,106	1,106	845	802	802
AICc	1,159	1,182	1,182	864	885	885
R ²	0.18	0.37	0.37	0.29	0.50	0.50
Adjusted R ²	(0.08)	(0.11)	(0.11)	0.04	0.07	0.07

Basic and robust GWR model achieve **higher R²/Adjusted R²** and **Lower AIC** value compared to Global model.

Insights:

- i. R² which measures goodness of fit is highest for Basic model for both cooked food and market stalls.
- ii. Lower AIC observed for Basic and Robust model for both cooked food and market stalls.
- iii. After compensating for number of variables in the model, Basic and Robust model still outperform Global model for cooked food and market stalls i.e. higher adjusted R².

GWR model perform **better** than linear regression mode

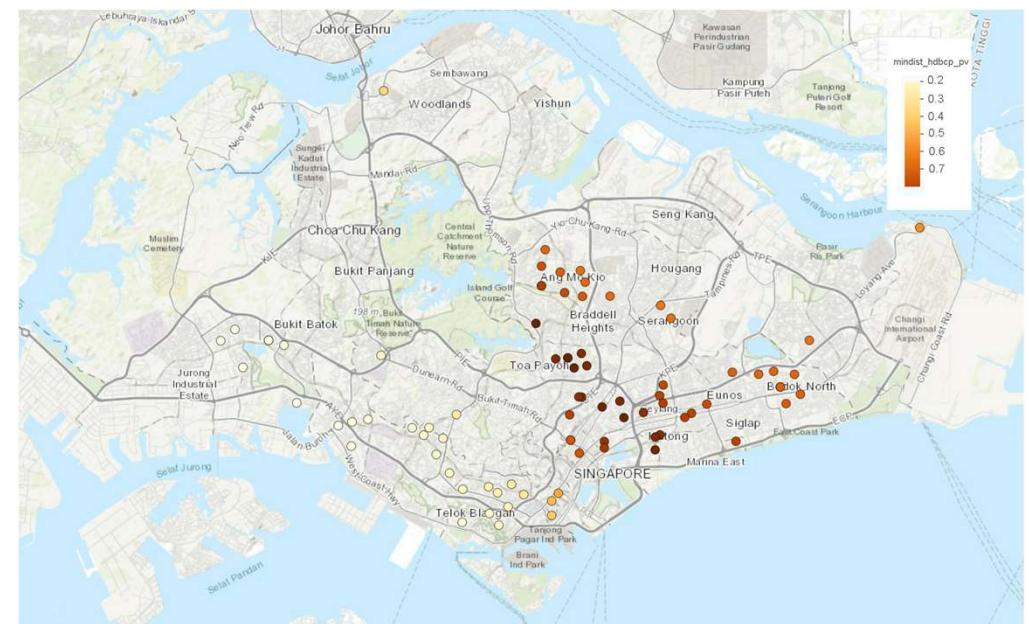
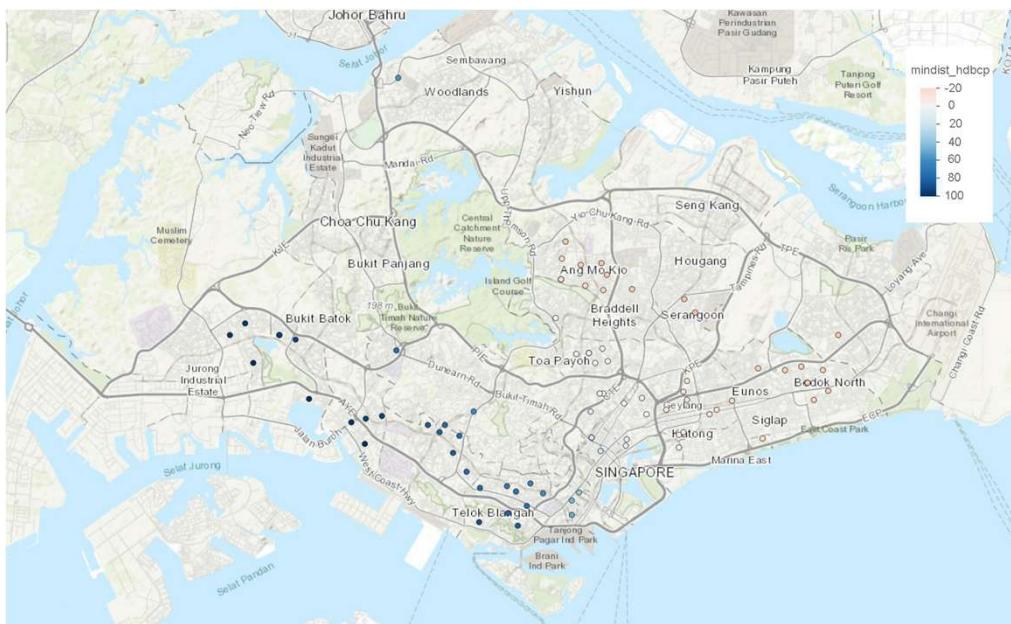
Interpretation of Results

Model Kernel	Cooked Food Stalls		Market Stalls	
	Basic / Robust (Filtered)	Bisquare	Basic / Robust (Filtered)	Bisquare
Optimal Bandwidth	88		81	
No. Observations	90		83	
Dependent Variable	Average Price		Median Price	
	Significant Variables			
	Variable	P value	Variable	P value
99.9% significance	No. of HDB carparks within 350m No. of MRT stations within 500m	0.000015 0.001024	No. of Bus Stops within 1km No. of HDB blocks within 1km Shortest Distance to Carpark	0.000631 0.000160 0.000000
99% significance			No. of HDB blocks within 750m	0.003601
90% significance	mindist_bstop bstop_0.5 bstop_1 hdb_0.5	0.027975 0.022614 0.017135 0.010448	Shortest Distance to MRT No. of MRT stations within 500m	0.046301 0.049937

- i. Cooked Food stalls depend largely on the availability of transport options, in the form of HDB carparks and MRT stations.
- ii. Market stalls are more dependent on a larger catchment area, and therefore the variables that are significant tend to be at a greater distance (1km).
- iii. Due to the nature of market stalls (produce, etc), the proximity of the nearest carpark is important to the price of

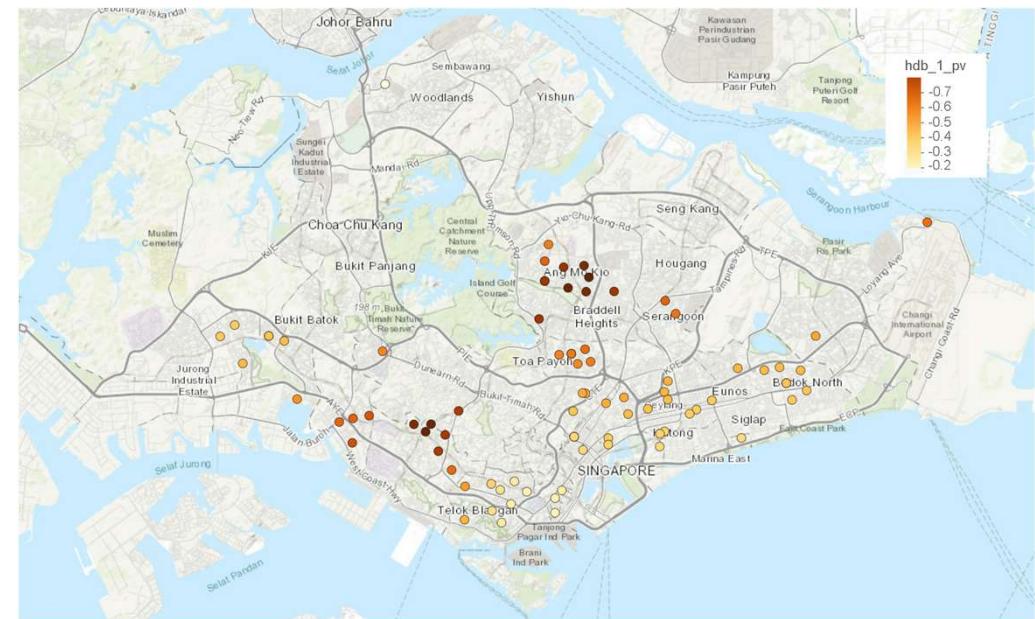
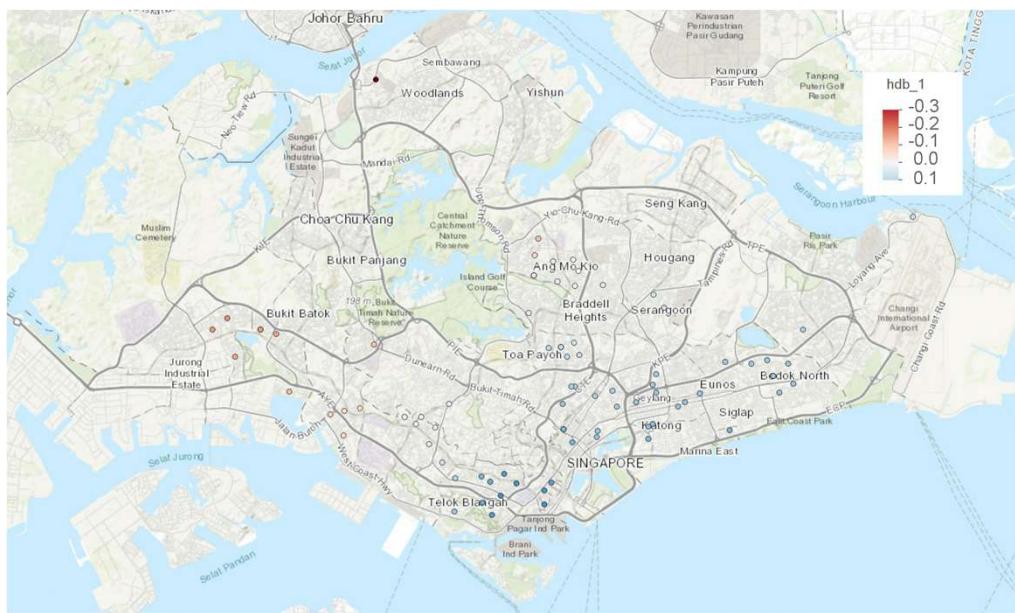
Most significant variables

Market Stalls – Minimum distance of **HDB carparks**



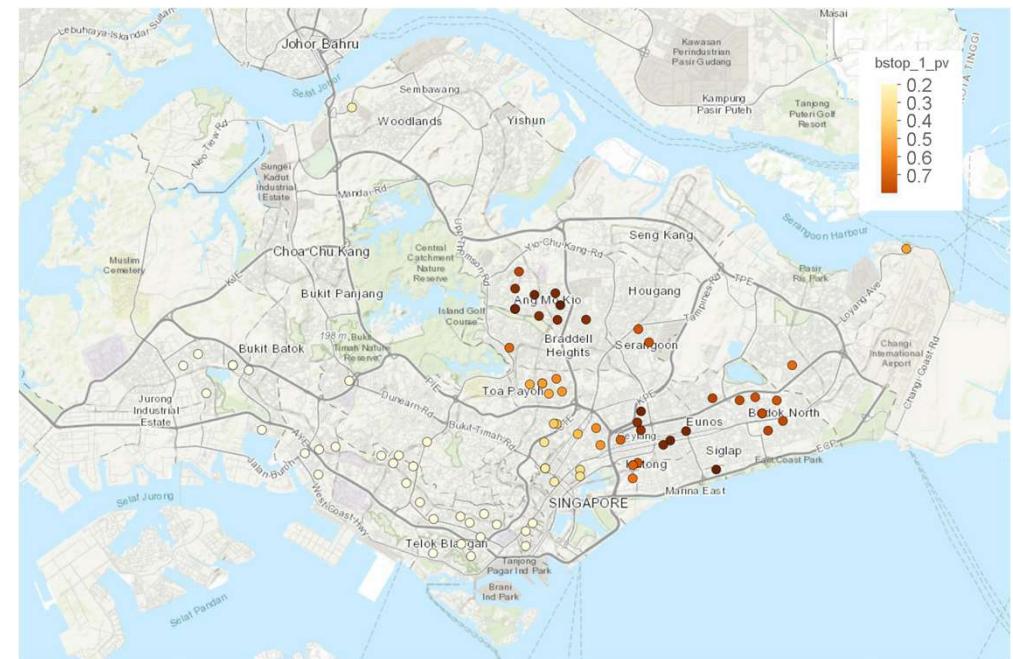
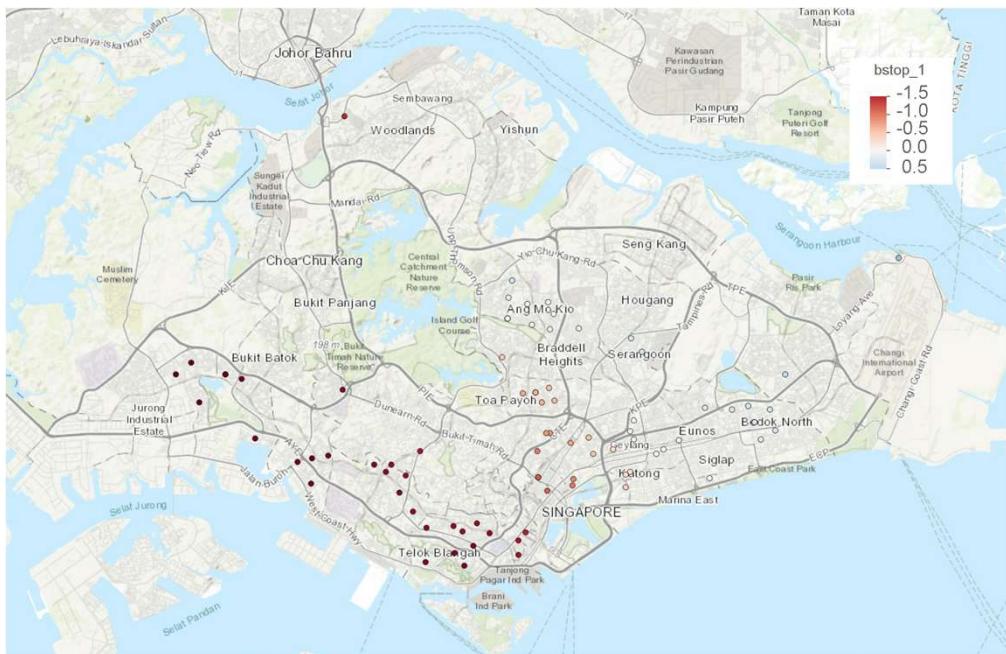
Most significant variables

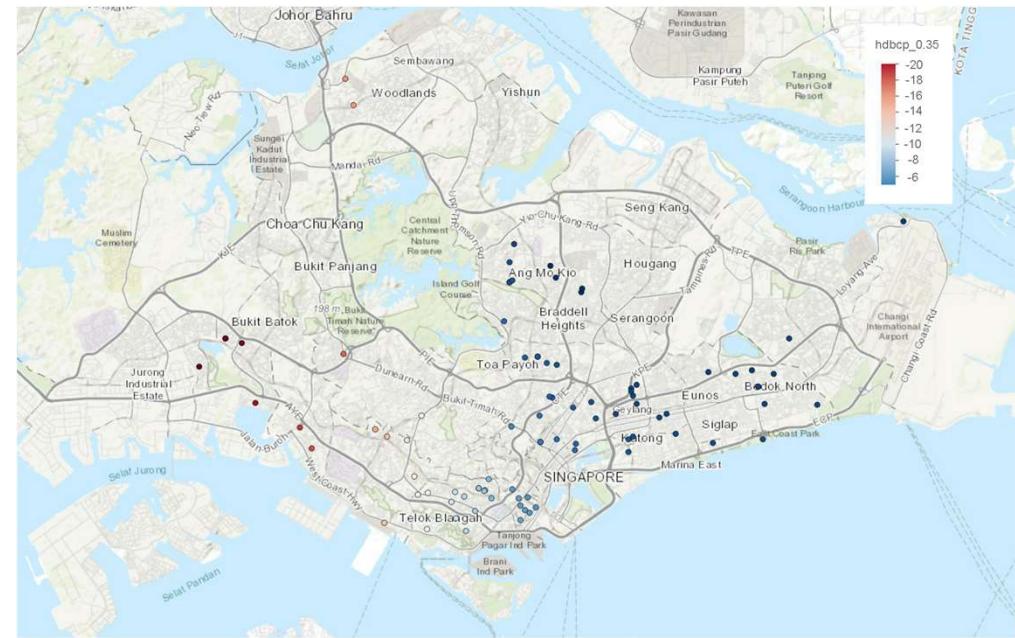
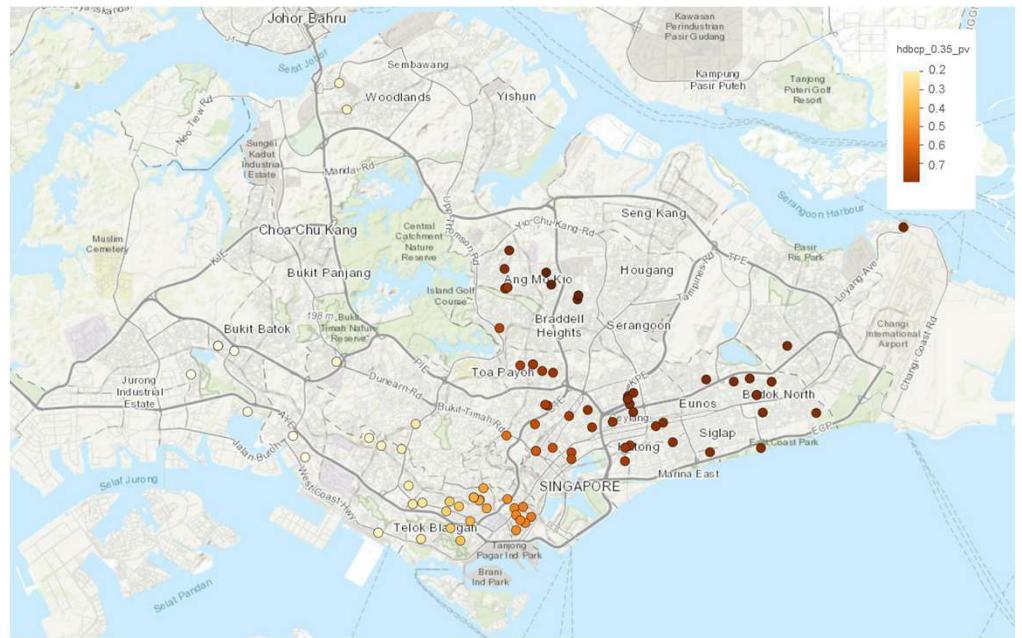
Market Stalls – Proximity to **HDB** within **1km**



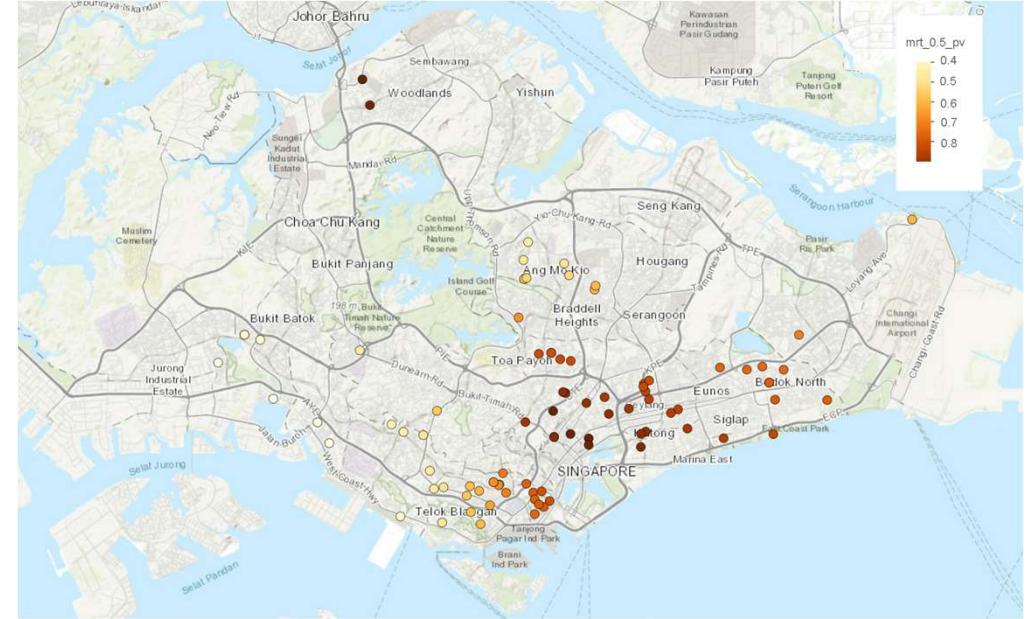
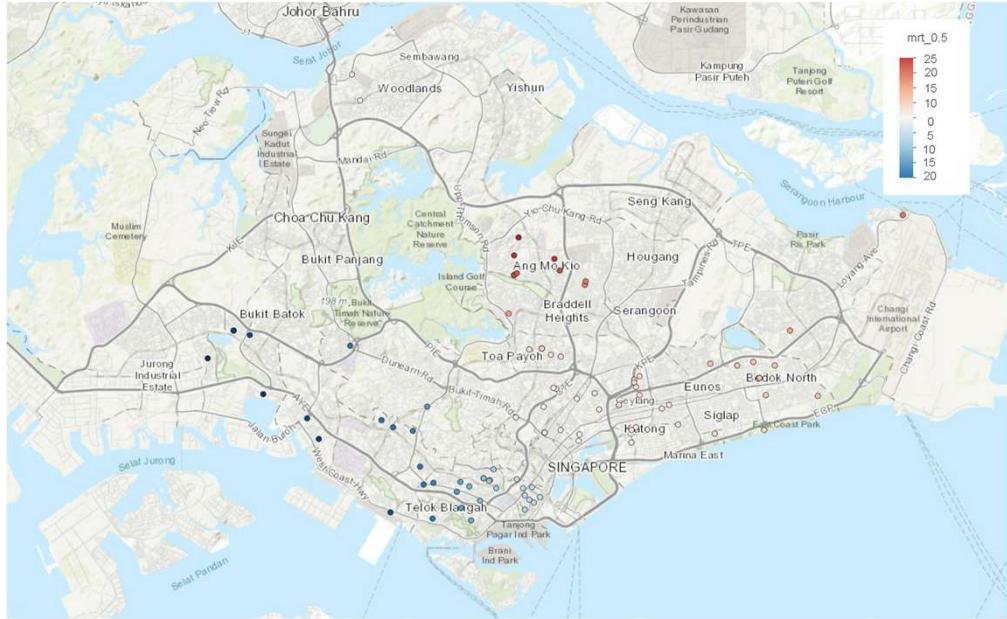
Most significant variables

Market Stalls – Proximity to **Bus Stops** within 1km





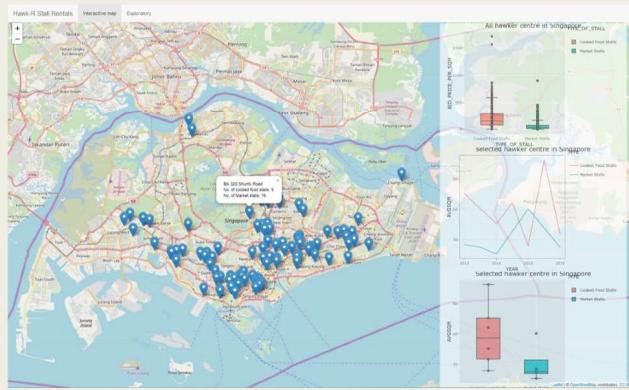
Most significant variables
Cooked Food Stalls – Proximity to **HDB** within 350m



Most significant variables
Cooked Food Stalls – Proximity to **MRT** within **500m**

Introduction

In Singapore, there are a total of 114 markets and hawker centres managed by the National Environmental Agency (NEA). Although information on past rental bids are publicly available on NEA's website, it is not in a form that is easily understandable to hawkers. Our group aims to build a shiny app to visualise Singapore's public hawker centre rental trends across Singapore, and specific hawker centre rentals. We have also included a geoweighted regression model for users to explore how locations and different variables affect the rental price of the hawker stalls.



Tools and Packages

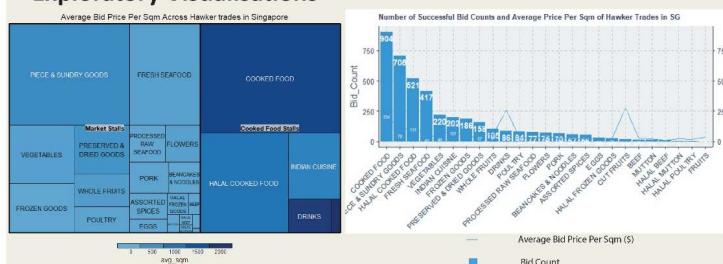


Data Preparation

Our project uses data from tender notices published on NEA's website as well as other publicly available information. Our final working file contains over 4,000 records of successful tender bids from Mar 2012 to Sep 2018. Locations information of key variables have been extracted from data.gov.sg and onestreetmap.

Column	Description	Example	Data Source
HAWKER_CENTRE	Name of hawker centre	Amoy Street Food Centre	NEA "List of 5 Highest Tender Bids"
STALL_NO	Unit number of stall	01-668	NEA "List of 5 Highest Tender Bids"
STALL_AREA	Size of stall (sqm)	5.65	NEA "Tender Bids for Hawker Stalls From March 2012 To September 2018"
TRADE	Specific type of trade	Vegetables	NEA "List of 5 Highest Tender Bids"
TRADE_GENERIC	Generic type of trade	Fresh Produce	Derived from "Trade"
TENDERED_BID	Successful bid price (\$)	\$2988	NEA "List of 5 Highest Tender Bids"
AVERAGE_BID_PRICE	Average successful bid price of Hawker Centre (\$)	\$450	Derived from "Tendered bid"
MONTBL	Month of successful bid	Apr	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
YEAR	Year of successful bid	2017	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
DATE	Month and Year of successful bid	Apr-2017	NEA "Tender Bids For Hawker Stalls From March 2012 To September 2018"
TYPE_OF_STALL	Type of business	Market Stalls	NEA
POSTAL_CODE	6 digits zip code representing postal location of hawker centre	560572	Derived from postal code using R
LAT	Latitude in 7.d.p.	1.2793399	Derived from postal code using R
LONG	Longitude in 7.d.p	103.8466525	Derived from postal code using R
BID_PRICE_PER_SQM	Tendered bid price of stall divided by stall size in sqm	\$28.85	Derived from tendered price and store size
LAST_UPGRADE	Reopening date after renovation	Feb-2003	NEA Hawker Centres Upgrading Programme
AGE_OF_HAWKER	Number of years since last renovation	15.8	Derived from fraction of year since reopening date after renovation
TYPE_OF_UPGRADING	Reason for upgrading	Reconfiguration	NEA Hawker Centres Upgrading Programme

Exploratory Visualisations



An interactive treemap shows the number of successful tender bids for each category of stalls (cooked food vs market stalls) and the sub-category breakdown. The average bid price psm is reflected as the intensity of colour.

GeoWeighted Regression (GWR) Model

Geographically Weighted (GW) models employ a moving window weighting technique where localised regression models are generated at target locations by weighing all neighboring observations found within a specific bandwidth.

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i$$

Results and Key Findings

Type of stall	Cooked food			Market			
	Bid price per SQM Model	Global	Basic	Robust	Global	Basic	Robust
AIC	1,142	1,106	1,106	845	802	802	
AICc	1,159	1,182	1,182	864	885	885	
R ²	0.18	0.37	0.37	0.29	0.50	0.50	
Adjusted R ²	(0.08)	(0.11)	(0.11)	0.04	0.07	0.07	

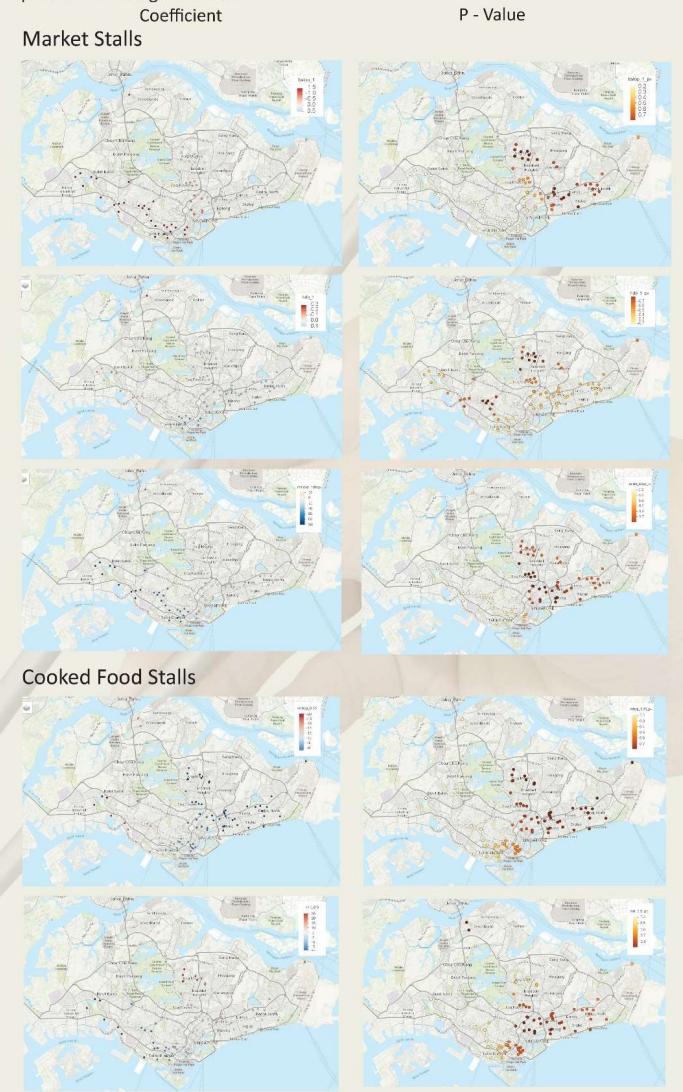
All GWR models perform better than linear regression model - r^2 value greater for basic and robust GWR model.

Model Kernel	Cooked Food Stalls		Market Stalls	
	Basic / Robust (Filtered)		Basic / Robust (Filtered)	
	Bisquare	Bisquare	Bisquare	Bisquare
Optimal Bandwidth	88		81	
No. Observations	90		83	
Dependent Variable	Average Price		Median Price	
99.9% significance	Significant Variables		Significant Variables	
	Variable	P value	Variable	P value
No. of HDB carparks within 350m	mindist_bstop	0.000015	No. of Bus Stops within 1km	0.000631
No. of MRT stations within 500m	bstop_0.5	0.001024	No. of HDB blocks within 1km	0.000160
	bstop_1		Shortest Distance to Carpark	0.000000
	hdb_0.5		No. of HDB blocks within 750m	0.003601
99% significance				
90% significance				
	mindist_bstop	0.027975	Shortest Distance to MRT	0.046301
bstop_0.5	0.022614	No. of MRT stations within 500m	0.049937	
bstop_1	0.017135			
hdb_0.5	0.010448			

- Cooked Food stalls depend largely on the availability of transport options, in the form of HDB carparks and MRT stations.
 - Market stalls are more dependent on a larger catchment area, and therefore the variables that are significant tend to be at a greater distance (1km).
 - However, due to the nature of market stalls (produce, etc), the proximity of the nearest carpark is important to the price of the stalls.
- These make sense intuitively and we can see how different hawker centres in Singapore are affected by these variables.

Coefficient and P-Value Plots

The most significant variables are plotted with the localised regression coefficients and p-values in the figures below:



Future Work

1. Current dataset has only ~4,000 *successful* tender bids for period of 2012-2018. We can explore a larger dataset, including non NEA-regulated ones, to be used, as well as all bid information.
2. Explore other variable inputs to drive better R^2 results.
3. To replicate project for appreciation of other retail formats e.g. HDB shophouses bid price.

References

Brunsdon, C. (2015, August 18). Geographically Weighted Regression. Retrieved from RPubs.

Brunsdon, C. (2015, Aug 10). Geographically Weighted Summary Statistics. Retrieved from RPubs.

Dennett, A. (2014, November 17). An Introduction to Geographically Weighted Regression in R. Retrieved from RPubs.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. JournalofStatisticalSoftware.

Lu, B. (2018, 08 20). Geographically-Weighted Models. Retrieved from Package GWmodel



Thank You

ISSS608 Visual Analytics Project Presentation

Group 1

Chia Yong Qing

Choo Mei Xuan

Clara Chua