

Museums in Singapore: Visitor Insights via Rshiny

Group Members:

Clarence TAY Cheng Wei

Heranshan S/O Subramaniam

LIU Chentao

Melissa TAN May Hua

Ranice TAN Hui Qi

Contents

1. Introduction	3
2. Objective	3
3. Data	3
3.1. Data Sources	3
3.2. Data Preparation	4
4. Descriptive Statistics	4
4.1. Visitorship Statistics	4
4.2. Visitor Demographic.....	5
4.3. Origin of Visitors	6
4.4. Repeat Visitorship	7
5. Inferential Statistics	8
5.1. Visitorship disruptions due to COVID-19 using Hypothesis Testing.....	8
5.2. Demographics association across different museums	10
5.3. Factors for repeat visitorship via Multilinear Regression and Logistic Regression Models	11
6. Conclusion	13
6.1. Inclusivity measured by demographics.....	13
6.2. Accessibility measured by location against population spread	14
6.3. Enjoyability measured by repeat visitorship	14
7. Recommendations	14
7.1. Crowd Management in an Endemic 'New Normal'	14
7.2. Capitalising on Local Tourism	14
7.3. Marketing for Greater Inclusivity	14
7.4. Accessibility Enhancements.....	15
8. Future Work	15
8.1. Understanding Group Visitors.....	15
8.2. Prediction of Repeat Visitorship	15
8.3. Using a more relevant benchmark year to assess post-COVID-19 norms.....	15

1. Introduction

Museums serve to preserve the heritage and culture of significance to a community and educate the public on the past, present and future. In Singapore, the National Heritage Board (NHB) preserves and protects Singapore's heritage. Its work includes the management of several national museums. The NHB aims to ensure that Singapore's heritage remains accessible to all citizens, and to promote greater appreciation of Singapore's heritage.

*"Our national museums and heritage institutions are cornerstones of the preservation and celebration of our heritage... This is why it is vital for our museums and institutions to continue being **accessible** to all, for all. Our shared goal is to make the Singapore museum experience **enjoyable** and **inclusive**... Ultimately, this ensures that our heritage remains a source of reflection and inspiration for us all¹."*

- National Heritage Board (NHB), Singapore

Based on the 2019 Heritage Awareness Survey, in 2018, 75% of Singaporeans said that a better understanding of Singapore's heritage and culture increases their sense of belonging to Singapore, an increase from 66% in 2014. The heritage participation of youths in particular, was high at 84%². The COVID-19 pandemic has also resulted in a rising trend of Singaporeans visiting our local museums as a form of local tourism³.

These trends point towards latent opportunities for increasing museum visitorship in Singapore, while achieving the NHB's mission of making museums accessible, enjoyable and inclusive for all.

2. Objective

The objective of this project is to conduct exploratory data analysis on NHB museum visitors over the past 3.5 years, to examine the extent to which museum visits have been (i) accessible, (ii) enjoyable and (iii) inclusive. The project also evaluates the impact of COVID-19 on museum visitorship.

An interactive R Shiny app was designed to allow members of the public to understand the visitor profile of each museum, as well as plan their visits to avoid crowded timings in a COVID-19 endemic 'new normal'. Today, there is no such interactive interface available to NHB.

For purposes of this study, the following proxies were used for the 3 parameters chosen for analysis:

- A. Inclusivity is defined as representativeness of museum visitors against national population by demographics (age, ethnicity, marital status)
- B. Accessibility is defined as the proximity of the location to public transport and subzones in Singapore.
- C. Enjoyability is defined as repeated visits (to the same or different museums)

3. Data

3.1. Data Sources

The data set was provided by the NHB for use in this study. It comprises data over 3.5 years from June 2017 to December 2020 and includes 6 national institutions, namely:

- i. National Museum Singapore (NMS)
- ii. Asian Civilisations Museum (ACM)
- iii. The Peranakan Museum (TPM)
- iv. Malay Heritage Centre (MHC)
- v. Indian Heritage Centre (IHC)
- vi. Sun Yat Sen Nanyang Memorial Hall (SYS)

¹<https://www.nhb.gov.sg/what-we-do/our-work/preserve-our-stories-treasures-and-places/museums-and-institutions>

²<https://www.nhb.gov.sg/what-we-do/our-work/sector-development/heritage-awareness-survey-2019>

³<https://www.straitstimes.com/life/arts/museum-attendance-on-the-rise-as-singaporeans-look-homeward-for-leisure>

By design, the data set excludes tourists and foreign visitors. No individuals were identifiable through the dataset provided. It has the following variables:

Data field	Description	How it will be used
MDAS ID	Unique identifier for each visitor	To identify repeat visitors
Date	Date of visit (dd/mm/YY)	To project peak/non-peak periods within a day, and throughout the year
Time	Time of visit	To project peak/non-peak periods within a day
Museum	Name of museum	To examine different visitorship and demographic patterns for each museum
Age	Age of visitor	For demographic analysis
Gender	Gender of visitor	For demographic analysis
Race	Race of visitor	For demographic analysis
Marital Status	Marital status of visitor	For demographic analysis
Region, Planning Area, Subzone	Residential location of visitor	To examine distribution of visitors' residential location (origin)

3.2. Data Preparation

- i. Missing data:
 - Data with missing demographic and residential locations was excluded (112,081 rows)
 - Data with missing visit time was excluded (20,932 rows)
- ii. Additional derived variables to facilitate analysis:
 - Repeat visitors were derived from MDAS ID field
 - Visitors' age was calculated from their year of birth and visit date
 - Museums' location was derived based on latitude and longitude data
 - Visitor Date, Time fields transformed into datetime objects in R using lubridate library
- iii. Additional datasets:
 - National population statistics from data.gov.sg were used to analyse the demographic patterns of museum visitors against national averages.

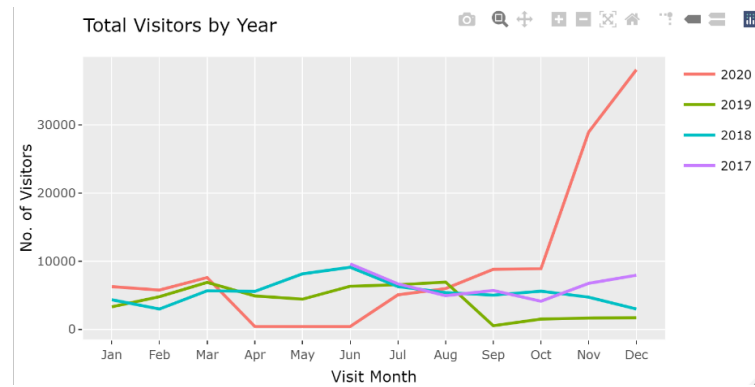
Other than "Shiny" and "Shiny Dashboard" for the development of R Shiny App, additional libraries were also used across our analysis, as below:

Data cleaning and transformation	readr, readxl, janitor, data.table, tidyverse, lubridate, reshape, dplyr, zoo
Visualisation	ggplot2, gplots, plotly, scales, corrplot
Geospatial mapping	leaflet, tmap, sf
Inferential analysis	ANOVA: shapiro.test, kruskal.test, pairwise.wilcox.test Chi-Square: chisq.test Regression: lm, glm

4. Descriptive Statistics

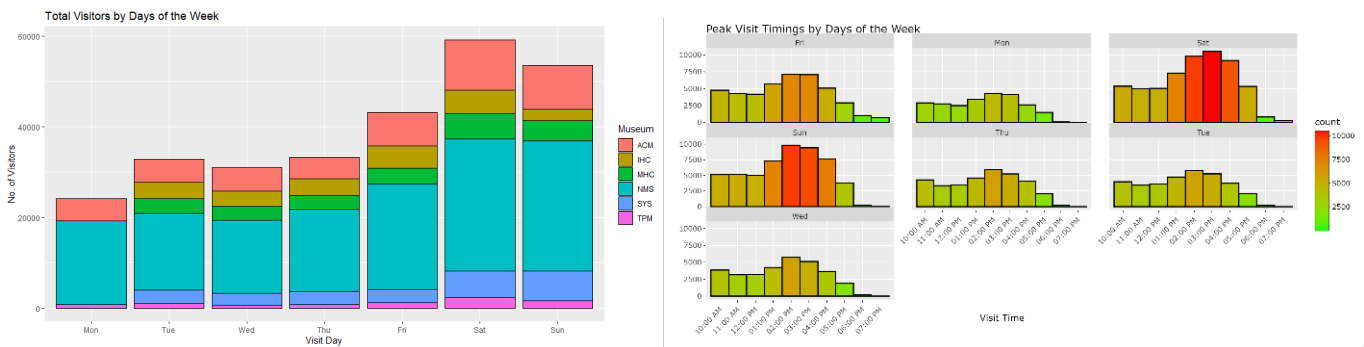
4.1. Visitorship Statistics

Visitorship to the various museums was grouped monthly and analysed. There were 277,351 unique visitors recorded in total. The month with the minimum and maximum number of visitors both occurred in 2020, with 0 in May, and 38,076 in December. This was most likely due to the impact of COVID-19 (More details in Section 5).



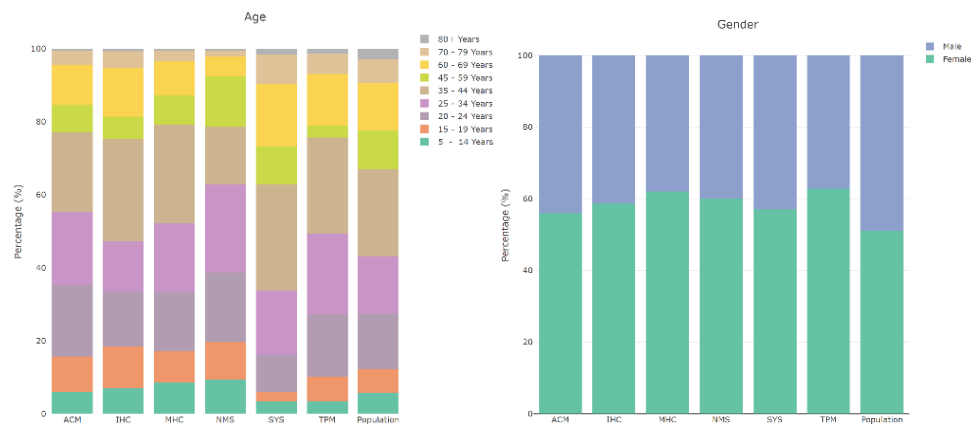
NMS welcomed the greatest number of visitors, 150,666 over the time period, and TPM had the fewest. This was expected given that TPM was closed since April 2019 for refurbishments. The community institutions, MHC, IHC and SYS, had similar visitor numbers averaging ~23,000 (Refer to Shiny app for details on each museum).

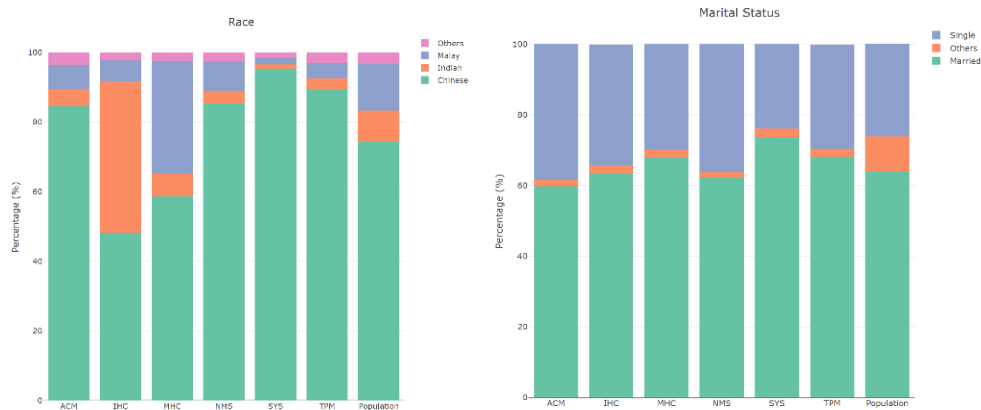
In terms of peak periods for museum visitors, the weekends were unsurprisingly the most popular days, with Saturday being the most crowded. Monday is most likely closure days for IHC, MHC and SYS as no visitors were recorded for these museums on that day. For any given day, the busiest time for all the museums was the early afternoon, from 1-3pm when visitors drop by the museums possibly after lunch. Visitors keen to avoid crowds should consider visiting in the evening, just before closing or just before lunch at around 11am.



4.2. Visitor Demographic

Visitor demographics was studied based on the composition of visitors to each museum across age, gender, race and marital status. These compositions were compared against the population demographics of Singapore to evaluate if there were any differences.

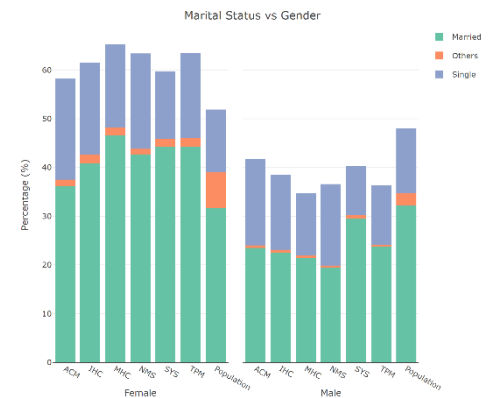




Based on the plot for age, all museums attracted a larger proportion of the younger generation (<45 years old) with most visitors being between 20 to 44 years old. SYS had the oldest demographic of visitors while NMS had the youngest.

Females also made up a higher proportion of visitors to all museums, higher than their population proportion. Except for IHC and MHC, the other museums attracted a higher proportion of Chinese residents compared to their population proportions. IHC, MHC and SYS, which are museums centered around the Indian, Malay and Chinese cultures saw an overwhelming proportion of visitors of those races.

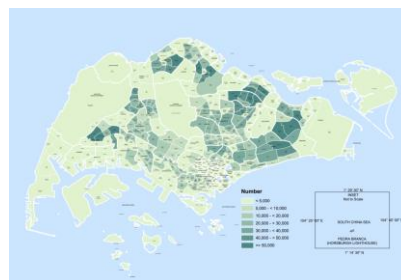
Race	ACM	IHC	MHC	NMS	SYS	TPM	Pop
Chinese	84.5	48.3	58.7	85.4	95.3	89.4	74.3
Indian	5.1	43.5	6.6	3.5	3.1	3.1	9.0
Malay	6.8	5.9	32.2	8.5	4.4	4.4	13.4
Others	3.7	2.3	2.4	2.7	3.1	3.1	3.2



Except for SYS, single visitors made up a larger proportion of visitors compared to population proportions. ACM and NMS saw a lower proportion of married visitors versus population proportions. Residents of other marital statuses (divorces and widowed) made up only a fraction of visitors (2.4% highest) though they have ~10% representation in the population. The chart below shows that married females made up a large proportion of visitors compared to their population proportion, which is the opposite of married males.

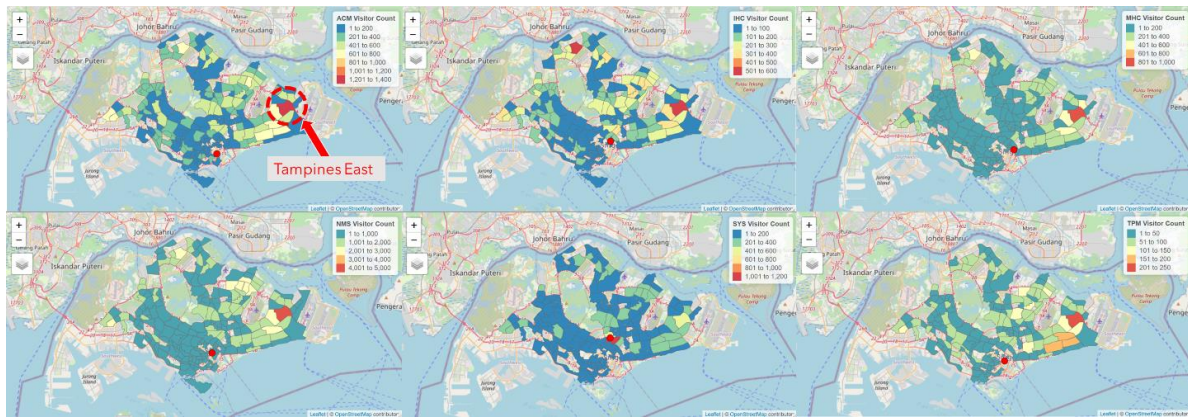
4.3. Origin of Visitors

The origin of visitors to the museums were studied via a choropleth. Map for total visitorship and individual museum visitor were generated, then compared against population across planning areas to identify the accessibility of the museums.



Source⁴

⁴ <https://www.singstat.gov.sg/-/media/files/publications/population/population2020.pdf>



From the choropleth, all museums are located in the Central region of Singapore, with ACM, NMS and TPM located close to the Central Business District of City Hall, IHC and MHC within the Rochor area, and SYS in Balestier.

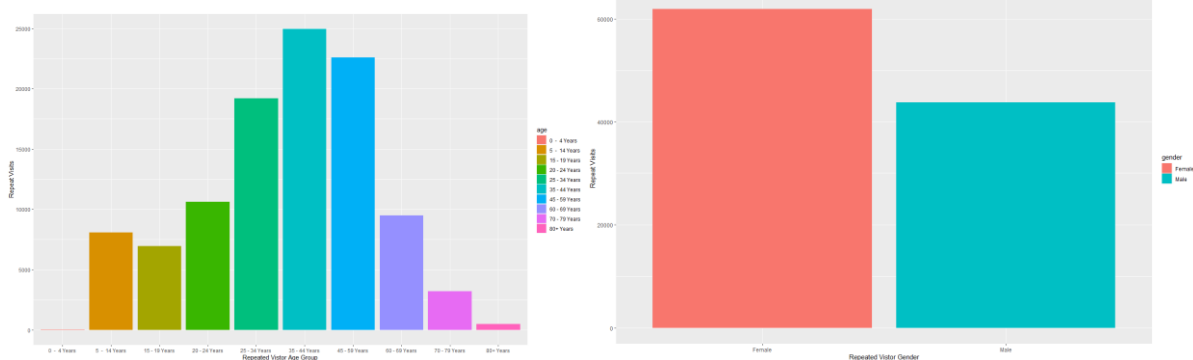
The residents in Tampines East seems to enjoy visiting museums, with most visitors come from there. There may higher arts and cultural museum enthusiasts in the area. IHC is also popular with residents staying in Woodlands East. Studying the public transportation network, shows that Tampines East and the other museums (except SYS) are located within 1.0 km of an MRT station on the Downtown Line. For SYS, most of their visitors originate from Balestier, which is where the museum is located.

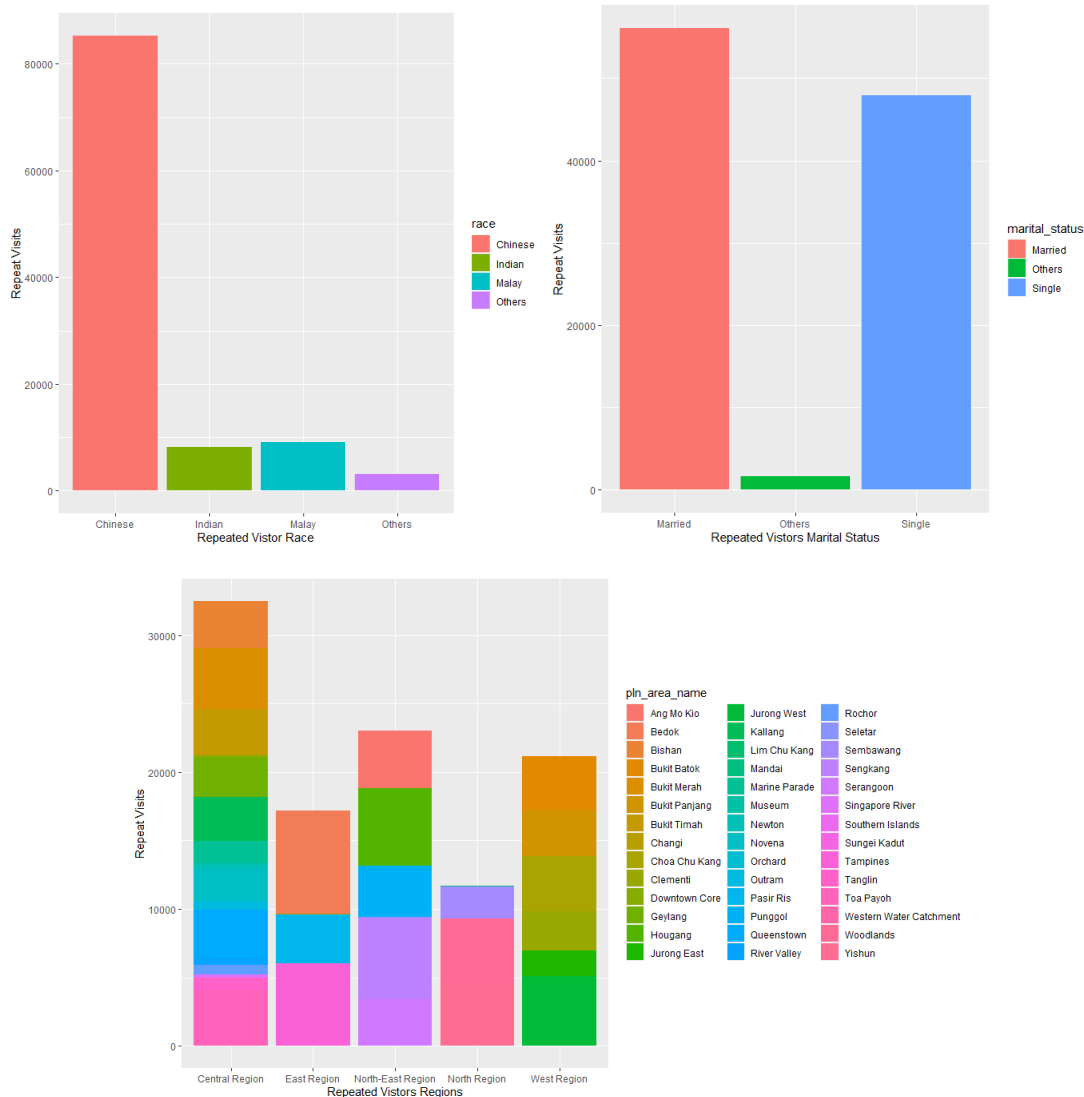
Populous towns like Jurong West, Yishun, Woodlands, Punggol, Seng Kang seem to contribute less than expected visitorship to the museum, compared to Tampines West. This may be attributed to the location of MRT stations in these towns, which are close to town centre whereas the location of Tampines West station's is within the heartlands. They may need to switch modes of transportation from heartlands to town centres, and then to museums.

For SYS, the lower visitorship even from Tampines East may be attributed to the location, which is located more than 2.0km from the Downtown Line Newton MRT station. It may also be a lesser-known museum, compared to the others.

4.4. Repeat Visitorship

We were also interested in the repeat visitors' demographics as we want to find out whether the demographics of repeat visitors of all museums demonstrate a specific pattern and whether such pattern is different from that of general visitorship.





Based on the graphs above, the age group 35-44 years have the highest number of repeat visitors; more repeat visitors are females than males; highest number of repeat visitors are of Chinese race, and lowest number of repeat visitors are from other races; highest number of repeat visitors are married, followed by repeat visitors who are single; the lowest number of repeat visitors have other marital status; the highest number of repeat visitors come from central region followed by the North-East Region, the West Region, East Region and finally North Region.

These observations of repeat visitors were similar to those of descriptive statistics of the general visitors.

5. Inferential Statistics

5.1. Visitorship disruptions due to COVID-19 using Hypothesis Testing

As mentioned in Section 4.1, there were observable disruptions to visitorship patterns in 2020 due to the impact of COVID-19. Two points of interest was studied, and hypothesis testing was conducted:

- Was the average number of visitors post-Circuit Breaker higher than before the Circuit Breaker, as more local tourists visited the museums following a period of physical distancing and the inability to travel abroad?

H₀: Average Visitors Post Circuit Breaker (June-December 2020) less than or equal to Average Visitors before Circuit Breaker (January 2018-March 2020) i.e. $\mu_x - \mu_y \leq 0$

H₁: Average Visitors Post Circuit Breaker (June-December 2020) greater than Average Visitors before Circuit Breaker (January 2018-March 2020) i.e. $\mu_x - \mu_y > 0$


```

> t.test(mu_x, mu_y,
+       alternative = "greater",
+       mu = 0, paired = FALSE, var.equal = FALSE,
+       conf.level = 0.90)

Welch Two Sample t-test

data: mu_x and mu_y
t = 1.6417, df = 6.0706, p-value = 0.07559
alternative hypothesis: true difference in means is greater than 0
90 percent confidence interval:
 1086.237      Inf
sample estimates:
mean of x mean of y
13741.429 4999.593

```

As the sample sizes for both means were <30 , and population variance was unknown, an upper-tail t-test was conducted at 90% confidence interval using R (see extract of codes above). The p-value was found to be 0.0756. Since this is smaller than the critical value of 0.1 at 90% confidence, there is sufficient evidence to reject the null hypothesis and infer that the average number of visitors post Circuit Breaker (June-December 2020) is greater than the average number of visitors before Circuit Breaker (January 2018-March 2020).

b) Was the average number of visitors pre-Circuit Breaker lower than the average number of visitors prior to COVID-19, as more Singaporeans would have had to adhere to safe distancing measures and/or would have been more cautious about visiting public spaces?

H₀: Average Visitors Pre-Circuit Breaker (January -March 2020) greater than or equal to Average Visitors before COVID-19 (January 2018-December 2019) i.e. $\mu_x - \mu_y \geq 0$

H₁: Average Visitors Pre-Circuit Breaker (January -March 2020) greater than Average Visitors before COVID-19 (January 2018-December 2019) i.e. $\mu_x - \mu_y < 0$

```

> t.test(mu_x, mu_y,
+       alternative = "less",
+       mu = 0, paired = FALSE, var.equal = FALSE,
+       conf.level = 0.90)

Welch Two Sample t-test

data: mu_x and mu_y
t = 2.4801, df = 5.2111, p-value = 0.9731
alternative hypothesis: true difference in means is less than 0
90 percent confidence interval:
 -Inf 2768.145
sample estimates:
mean of x mean of y
6545.667 4806.333

```

As the sample sizes for both means were <30 , and population variance was unknown, a lower-tail t-test was conducted at 90% confidence interval using R. The p-value was found to be 0.973. Since this is larger than the critical value of 0.1 at 90% confidence, there is insufficient evidence to reject the null hypothesis. Hence, we are unable to conclude that the average number of visitors pre-Circuit Breaker (January-March 2020) was lower than the average number of visitors before COVID-19 (January 2018-December 2019).

5.2. Demographics association across different museums

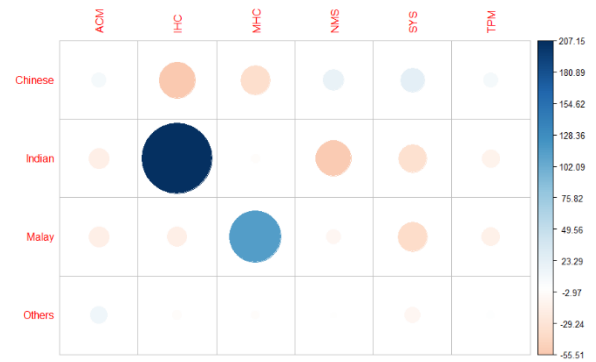
a) Is race a factor in the choice of museum visited?

H₀: No association exists between race and museum visited

H₁: Association exists between race and museum visited

Pearson's Chi-squared test

data: df_race
X-squared = 68802, df = 15, p-value < 2.2e-16



Chi-squared test was conducted based on the composition of race of visitors to the different museums. At a confidence level of 90%, the test yielded a p-value of <2.2e-16. Since this is <0.1, we have enough evidence to conclude at a 90% confidence level that the preference for museums is not independent of race. The plot for Pearson residuals is shown below. The Indian community has a strong positive association with IHC and a moderate negative association with NMS. The Malay community has a strong positive association with MHC and a moderate negative association with SYS. The Chinese community has a weak positive association with SYS and a moderate negative association with IHC and MHC.

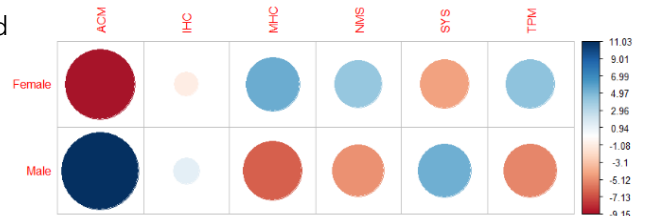
b) Does gender influence the choice of museum visited?

H₀: No association exists between gender and museum visited

H₁: Association exists between gender and museum visited

Pearson's Chi-squared test

data: df_gender
X-squared = 419.94, df = 5, p-value < 2.2e-16



Chi-squared test was conducted based on the composition of race of visitors to the different museums. At a confidence level of 90%, the test yielded a p-value of <2.2e-16. Since this is <0.1, we have enough evidence to conclude at a 90% confidence level that an association exists between gender and museum visited. The plot for Pearson residuals is shown below. Females have a moderate positive association with MHC and a strong negative association with ACM. Males had the opposite association.

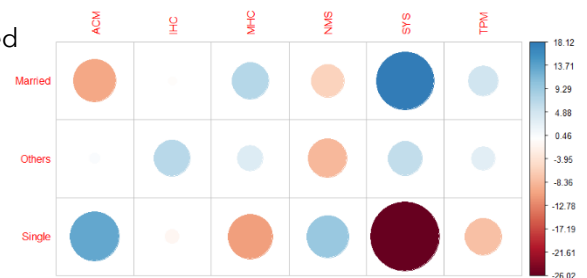
c) Marital status influences the choice of museum visited?

H₀: No association exists between marital status and museum visited

H₁: Association exists between marital status and museum visited

Pearson's Chi-squared test

data: df_ms
X-squared = 1857.7, df = 10, p-value < 2.2e-16



Chi-squared test was conducted based on the composition of race of visitors to the different museums. At a confidence level of 90%, the test yielded a p-value of <2.2e-16. Since this is <0.1, we have enough evidence to conclude at a 90% confidence level that an association exists between museums and marital status. The plot for Pearson residuals is shown below. Married visitors showed a strong positive association with SYS and a moderate

negative association with ACM. The opposite was true for singles. Those of other marital status had a weak positive association with IHC and SYS, while a weak negative association existed with NMS.

d) Does age influence the choice of museum visited?

H₀: There is no difference in mean age across museums

H₁: There is a difference in mean age across museums

```
> leveneTest(Age ~ Museum, data = df_age_anova)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  5  542.86 < 2.2e-16 ***
      277345
```

Museum	Mean Age	Std Deviation
ACM	38.9	17.5
IHC	41.3	18.2
MHC	38.9	17.3
NMS	33.4	16.2
SYS	45.0	19.5
TPM	44.0	16.8

The p-value for Levene's test is less than the significance level of 0.1. This means that there is evidence to suggest that the variance across groups is statistically significant at a 90% confidence interval. The data fails the homogeneity of variance assumption required for ANOVA testing. Therefore, the [Kruskal-Wallis test](#) was conducted.

```
Kruskal-wallis rank sum test
data: Age by Museum
Kruskal-Wallis chi-squared = 14618, df = 5, p-value < 2.2e-16
```

	ACM	IHC	MHC	NMS	SYS
IHC	<2e-16	-	-	-	-
MHC	0.26	<2e-16	-	-	-
NMS	<2e-16	<2e-16	<2e-16	-	-
SYS	<2e-16	<2e-16	<2e-16	<2e-16	-
TPM	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

p value adjustment method: BH

As the p-value is less than the significance level 0.1, we can conclude that there are significant differences between the museums. A [pairwise Wilcox test](#) was conducted to identify which pairs of museums are different. Output is as shown above (right). The pairwise comparison table on the right shows that except for MHC-ACM, the mean age for the other pairs were all significantly different ($p < 0.1$).

5.3. Factors for repeat visitorship via Multilinear Regression and Logistic Regression Models

Upon examining the demographics of repeated visitorship, we wanted to see whether certain demographics of visitors correlate to a higher repeat visit. We generated a multiple linear regression model with age, gender, race, marital status, region name, subzone name as independent variables and number of repeat visits as dependent variable y.

We deemed the variable birth month as irrelevant and excluded it from the x variables.

We first constructed the multi-linear regression model manually and then built the same multi-linear regression models with stepwise selection to identify which variables were insignificant and could be eliminated. Below is the output of the stepwise regression.

Although genderMale and marital_statusSingle turned out to be most significant, the overall R square value (0.01082) was very low which indicates an absence of a linear relationship, or that the selected independent variables may not be sufficient to explain the variation of dependent variable y (repeat visits).

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.74911    0.43508   4.020 5.83e-05 ***
age5 - 14 Years    0.33941    0.43508   0.780 0.43533
age15 - 19 Years    0.41716    0.43534   0.958 0.33794
age20 - 24 Years    0.55842    0.43490   1.284 0.19914
age25 - 34 Years    0.69947    0.43474   1.609 0.10763
age35 - 44 Years    0.87700    0.43496   2.016 0.04378 *
age45 - 59 Years    0.97137    0.43493   2.233 0.02553 *
age60 - 69 Years    1.05222    0.43563   2.415 0.01572 *
age70 - 79 Years    0.96234    0.43775   2.198 0.02793 *
age80+ Years       0.89532    0.45500   1.968 0.04911 *
genderMale         0.11721    0.01881   6.232 4.66e-10 ***
raceIndian         0.03481    0.03391   1.026 0.30473
raceMalay        -0.01307    0.03263  -0.401 0.68873
raceOthers         0.12849    0.05460   2.353 0.01862 *
marital_statusothers -0.01964    0.07435  -0.264 0.79163
marital_statussingle 0.16800    0.02451   6.854 7.28e-12 ***
region_nameEast Region -0.05730    0.02757  -2.078 0.03772 *
region_nameNorth-East Region -0.04542    0.02511  -1.809 0.07047 .
region_nameNorth Region -0.04889    0.03162  -1.546 0.12203
region_nameWest Region -0.06664    0.02580  -2.583 0.00981 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.789 on 40494 degrees of freedom
Multiple R-squared:  0.01128,    Adjusted R-squared:  0.01082
F-statistic: 24.32 on 19 and 40494 DF,  p-value: < 2.2e-16

```

Next, a Logistic Regression was performed to investigate if there was any association between the demographic variables and the likelihood of a repeat visit. Repeat visitors were assigned to the response '1' while non-repeat visitors were assigned to '0':

```

Call:
glm(formula = `Repeat Visitor 2` ~ Age + Gender + Race + `Marital Status`,
     family = binomial(link = logit), data = my_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2291 -0.9797 -0.9287  1.3694  1.6178

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.8579900  0.0151594 -56.598 < 2e-16 ***
Age            0.0083461  0.0002892  28.861 < 2e-16 ***
GenderMale     0.0505188  0.0080631   6.263 3.72e-10 ***
RaceIndian     0.1208867  0.0150617   8.028 1.01e-15 ***
RaceMalay     -0.0538067  0.0139821  -3.848 0.000119 ***
RaceOthers     0.1053757  0.0239315   4.403 1.07e-05 ***
`Marital Status`Others -0.1903151  0.0319117  -5.964 2.46e-09 ***
`Marital Status`single  0.0812639  0.0102452   7.932 2.16e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 368553  on 277350  degrees of freedom
Residual deviance: 367355  on 277343  degrees of freedom
AIC: 367371

Number of Fisher Scoring iterations: 4

```

Using Gender as an example, the output above was interpreted as follows:

H₀: There is no association between Repeat Visitorship and Gender when controlling for Age, Race and Marital Status

H₁: There is an association between Repeat Visitorship and Gender when controlling for Age, Race and Marital Status

Since the p-value for GenderMale is very small and less than the critical value of 0.1 at 90% CI, we have sufficient statistical evidence to reject H₀ and infer that there is an association between Repeat Visitorship and being Male

Further, to summarise the results of the logistic regression, being Male, Indian/Others, and/or Single is positively correlated with being a Repeat Visitor, as observed by the positive coefficient estimates, while being Malay, and/or Widowed/Divorced are negatively correlated with being a repeat visitor, as observed by the negative coefficient estimates.

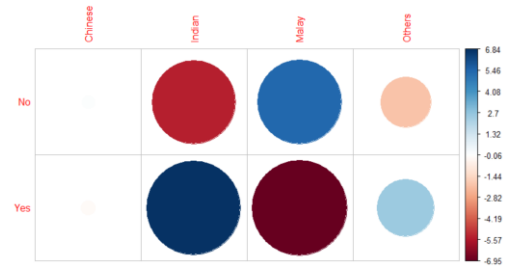
Additionally, chi-square tests were conducted to further understand the strength of association between race, gender and marital status on repeat visits.

H₀: No association exists between race and repeat visits
H₁: Association exists between race and repeat visits

Race

Pearson's Chi-squared test

data: df_race
X-squared = 164.13, df = 3, p-value < 2.2e-16

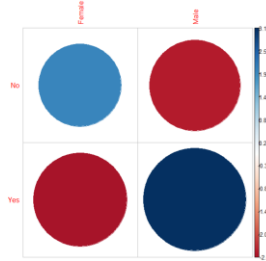


H₀: No association exists between gender and repeat visits
H₁: Association exists between gender and repeat visits

Gender

Pearson's Chi-squared test

data: df_gender
X-squared = 26.74, df = 1, p-value = 2.328e-07

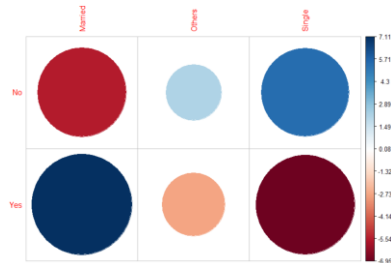


H₀: No association exists between marital status and repeat visits
H₁: Association exists between marital status and repeat visits

Marital Status

Pearson's Chi-squared test

data: df_ms
X-squared = 172.27, df = 2, p-value < 2.2e-16



The p-value for all three demographics were smaller than alpha at 90% confidence interval, indicating that there is sufficient evidence to reject the null hypothesis. As such it can be inferred that race, gender and marital status influenced if an individual would be a repeat visitor. For race, there was a strong positive association between Indian visitors and repeat visits, while a strong negative association existed between Malay visitors and repeat visits. For gender, a strong positive association was seen between male visitors and repeat visits, while the opposite was true for female visitors. For marital status, married visitors had a strong positive association with repeat visits while the opposite was true for single visitors. Visitors of other marital status had a negative association with repeat visits.

6. Conclusion

Overall, the demographic profile of visitors to all the museums could be more inclusive, and align more closely to that of the national population. Accessibility can be enhanced for residents staying in the Western and Northern parts of the island. Enjoyability of the museum experience could be improved for the Malay community, as well as those who were Widowed/Divorced.

6.1. Inclusivity measured by demographics

Age: Museums in general attracted a younger demographic compared to the general population. SYS had the highest mean age of visitors while NMS had the lowest. More could be done to make museums accessible to the older demographics.

Gender: Visitors were also largely female. These female visitors had a greater affinity to museums such as MHC, NMS and TPM. They had a strong negative correlation to ACM and a mild negative correlation with SYS. The opposite was true for male visitors, who made up a smaller proportion of visitors compared to their population proportions. Male visitors were strongly associated with ACM and mildly associated with SYS. They have a greater affinity to MHC, NMS and TPM.

Marital Status: Residents of other marital status were underrepresented in the visitor profiles of all museums. Except for SYS, singles were overrepresented in the visitor profiles of museums, compared to their population proportions. Single visitors had a strong negative correlation to SYS and a strong positive correlation to ACM. Married females contributed above their population proportions to museum visits. However, the opposite was true for married males. Married visitors had the opposite correlation of the single visitors.

Race: Except for IHC and MHC, members of the minority ethnic groups were underrepresented in all museums. For IHC, MHC and SYS, which are community centric museums, there was an overrepresentation of the ethnic community they are centered on. Among non-community centric museums, the residents of Indian ethnicity showed the strongest negative correlation NMS.

6.2. Accessibility measured by location against population spread

With the exception of SYS, the museum visitor choropleth shows that they are generally popular with the residents staying in Tampines East. This may be because the residents enjoy visiting museums, or better accessibility from Tampines East to the museums, which are located within 1.0km of a Downtown Line station.

The low visitorship on SYS museum from potential museums enthusiasts, but high visitorship from residents staying near the museum may also highlight the location constraint of the museum, which is more than a 10 minutes' walk from the closest MRT station. It may also be a lesser-known museum. Hence, enhancing accessibility and awareness of SYS museum may be tools employed to increase visitor traffic.

Areas with higher population like Jurong West and Punggol also seem to contribute less than proportionate visitorship to the museum. This may be due to the location of their MRT stations, which are located in the individual town centres, whereas Tampines East MRT is located within the heartlands, reducing the need to switch transportation modes. Alternatively, Community efforts through People Association or Community Centers may be used to enhance traffic to museums in such regions with disproportionately low visitorship.

6.3. Enjoyability measured by repeat visitorship

In general, the museums have a similar percentage of repeat visitors. The observations and demographics of repeat visitors are also like those of the general visitors. The results of the logistic regression suggests that more could be done to encourage repeat visitorship by those from the Malay community, as well as those who are widowed and divorced.

7. Recommendations

7.1. Crowd Management in an Endemic 'New Normal'

The analysis showed that visitors had clear preferences for visit days and timings. To spread out the numbers on peak days/times, the museums may wish to consider adapting the Shiny app for use on their websites so that visitors can conveniently check and decide if they wished to visit a different, less crowded museum instead or change the timing of their visit.

7.2. Capitalising on Local Tourism

The average number of visitors post-Circuit Breaker (June-December 2020) was found to be statistically greater than the average number of visitors before. The NHB should not miss this opportunity when local tourism is on an uptrend, to sustain this increased level of visitorship, while attracting more who have discretionary time now that options for travel abroad are still relatively limited.

7.3. Marketing for Greater Inclusivity

First, museums could offer incentives to visitors to stimulate repeat visits (4.4 Repeat Visitorship), especially to those segments with a negative association. For example, the museums could offer museum-related merchandise to visitors after certain number of return visits.

Specifically, museums could encourage overrepresented demographics (e.g. Married females) to bring their families member (e.g. Married males) to increase visits of underrepresented demographics to museums (4.2 Visitor demographics). To achieve these, museums could hold more group or family-friendly events or exhibitions when COVID-19 movement measures allow.

Furthermore, although community museums tend to attract members within their own communities, by introducing new, cross-cultural content, they could attract members of other ethnic communities (5.2 Demographics association across different museums).

Lastly, highly visited museums can collaborate with museums that gets fewer visits. For instance, NMS could collaborate with TPM as the former gets the highly number of visits and the latter gets the least number of visits (4.1 Visitor Statistics).

7.4. Accessibility Enhancements

Towns such as Jurong West and Yishun, which contribute less than expected visitorship based on the population by subzone were identified. A potential cause may be the location of MRT stations which are close to the city centre instead of heartlands. To increase accessibility for these residents, museums may want to work with community partners such as People's Association to organize trips or plan exhibitions showcases in the town centers.

The analysis identified SYS as a museum potentially facing accessibility constraints, due to its remoteness from high traffic public transportations like MRT. To increase visitorship, the museum may consider implementing shuttle services from nearby MRT stations.

8. Future Work

8.1. Understanding Group Visitors

Given the data available, this study examined visitorship as an individual activity only. To better assess accessibility and enjoyability, it may be useful to find a way to tag visitor ID to the social groups that they were in during the visit (e.g. as a family, a couple, school trip or local tour group).

8.2. Prediction of Repeat Visitorship

Other important variables could have been left out from the multilinear regression model for repeat visitors due to a lack of data or incomplete data. These include variables that could potentially be better predictors than demographics e.g. income levels of visitors, the type of exhibition attended, and whether the ticket was paid or unpaid. A predictive model could then be developed for repeat visitorship, extending from the logistic regression performed.

8.3. Using a more relevant benchmark year to assess post-COVID-19 norms

Finally, while the dataset provided a useful view of the impact of the early stages of the impact of COVID-19 in 2020, to better assess visitorship patterns in the endemic phase, it may be better to repeat the study with data from 2022 onwards.