Amenity index

Yuxuan

18/05/2021

Import libraries

```
library(dplyr)
library(readr)
library(ggplot2)
library(tidyr)
library(imputeTS)
library(qwraps2)
options(qwraps2_markup='markdown')
library(hablar)
```

Import data

```
review_poi<- read_csv("~/Desktop/MDS/data599/google-reviews-arts/google_reviews_poi_with_hours.csv")
van_poi<-read_csv("~/Desktop/MDS/data599/w2020-data599-capstone-projects-statistics-canada-transit/data/
```

Merge review dataset with vancouver point of interest

```
left_join(review_poi,van_poi,by=c("poi_name"="name"))%>%distinct()->merged_data
```

Convert the data to numeric

```
merged_data%>% convert(num(Rating, Total_Review,open_days,Total_hours))->merged_data
```

```
## Warning in as_reliable_num(.): NAs introduced by coercion
```

Number of amenity in each type of arts facility

```
merged_data%>%group_by(type)%>%count()
```

```
## # A tibble: 9 x 2
## # Groups:   type [9]
##   type                                n
##   <chr>                           <int>
## 1 art or cultural centre              5
## 2 artist                             48
## 3 festival site                       2
## 4 gallery                            99
## 5 heritage or historic site          28
## 6 library or archives                86
## 7 miscellaneous                       6
## 8 museum                             92
## 9 theatre/performance and concert hall    75
```

EDA on museum

```
merged_data%>%filter(type=="museum")%>%arrange(desc(Total_Review))%>%distinct()->poi_museum
poi_museum%>%select(poi_name,open_days,Total_hours,Rating,Total_Review)->poi_museum
```

Find the percentage of NAs in poi_museum

```
poi_museum[poi_museum == 0] <- NA
colMeans(is.na(poi_museum))
```

```
##     poi_name    open_days  Total_hours       Rating Total_Review
##   0.00000000   0.29347826   0.29347826   0.06521739   0.06521739
```

Fill na value with its mean

```
poi_museum<-na_mean(poi_museum)
```

museum summary table

```
summary<- list(
  "Rating"=list(
    "min"= ~ min(Rating,na.rm = TRUE),
    "max"= ~ max(Rating,na.rm = TRUE),
    "mean"= ~ mean(Rating,na.rm = TRUE)),

  "Total_Review"=list(
    "min"= ~ min(Total_Review,na.rm = TRUE),
    "max"= ~ max(Total_Review,na.rm = TRUE),
    "mean"= ~ mean(Total_Review,na.rm = TRUE)),

  "Total_hours"=list(
    "min"= ~ min(Total_hours,na.rm = TRUE),
    "max"= ~ max(Total_hours,na.rm = TRUE),
    "standard deviation"= ~ sd(Total_hours,na.rm = TRUE),
    "mean"= ~ mean(Total_hours,na.rm = TRUE)),

  "Open_days"=list(
    "min"= ~ min(open_days,na.rm = TRUE),
    "max"= ~ max(open_days,na.rm = TRUE),
     "standard deviation"= ~ sd(open_days,na.rm = TRUE),
    "mean"= ~ mean(open_days,na.rm = TRUE))



)


whole<-summary_table(poi_museum,summary)
whole
```

	poi_museum (N = 92)
Rating	
min	3.4
max	5
mean	4.43023255813953
Total_Review	
min	1
max	8833

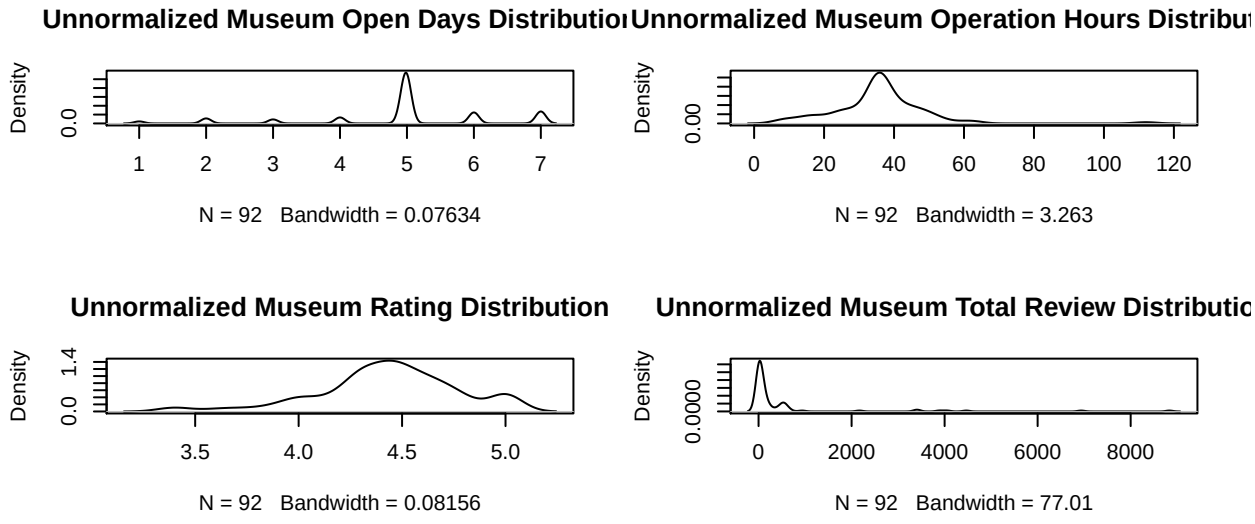
	poi_museum (N = 92)
mean	529.244186046512
Total_hours	
min	8
max	112
standard deviation	13.4659270711317
mean	35.8807692307692
Open_days	
min	1
max	7
standard deviation	1.31741697629628
mean	4.96923076923077

```
summary(poi_museum)
```

poi_name open_days Total_hours Rating
Length:92 Min. :1.000 Min. : 8.00 Min. :3.40
Class :character 1st Qu.:4.969 1st Qu.: 30.00 1st Qu.:4.30
Mode :character Median :5.000 Median : 35.88 Median :4.43
Mean :4.969 Mean : 35.88 Mean :4.43
3rd Qu.:5.250 3rd Qu.: 42.00 3rd Qu.:4.60
Max. :7.000 Max. :112.00 Max. :5.00
Total_Review
Min. : 1.0
1st Qu.: 9.5
Median : 41.5
Mean : 529.2
3rd Qu.: 292.8
Max. :8833.0

Unnormalized density plot of museum

```
par(mfrow = c(3, 2))
plot(density(unlist(poi_museum[,2])), main = 'Unnormalized Museum Open Days Distribution')
plot(density(unlist(poi_museum[,3])), main = 'Unnormalized Museum Operation Hours Distribution')
plot(density(unlist(poi_museum[,4])), main = 'Unnormalized Museum Rating Distribution')
plot(density(unlist(poi_museum[,5])), main = 'Unnormalized Museum Total Review Distribution')
```



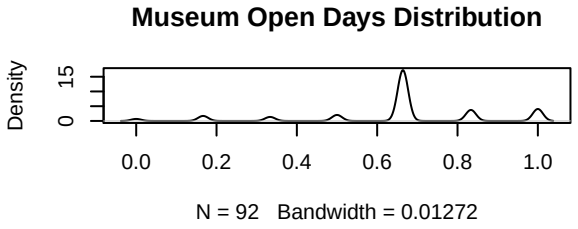
Normalization/ Min-Max scaling

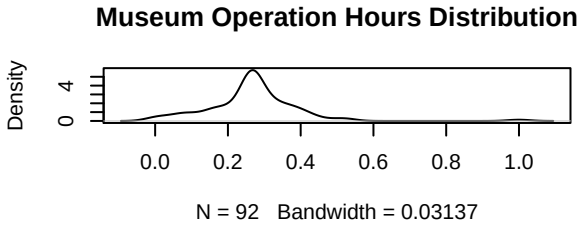
$$Transformed.Values = \frac{Values - Mean}{Max - Min}$$

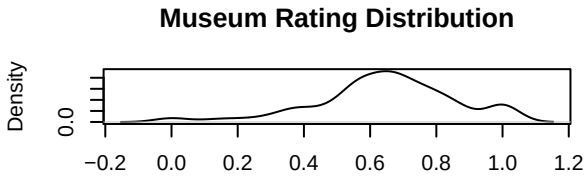
```
#################################
## Computing Museum Weight Index
#################################
poi_museum$Rating<-as.numeric(poi_museum$Rating)
normalize <- function(x) {
return ((x - min(x)) / (max(x) - min(x)))
}
Norm_museum<-poi_museum%>%na.omit()%>% mutate_if(is.numeric, normalize)
```

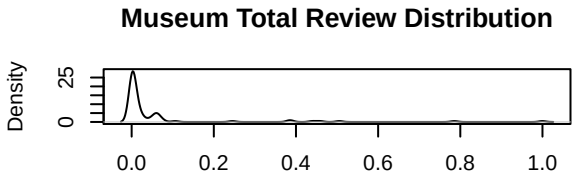
```
# Navie weighted index
Norm_museum%>%mutate(Index=(open_days+Total_hours+Rating+Total_Review)/4)->Norm_museum
```

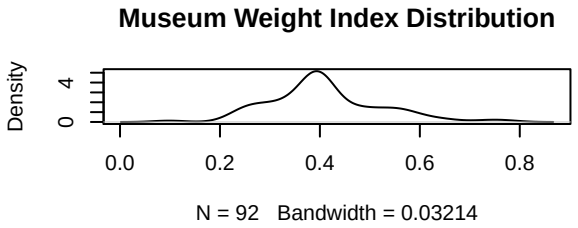
```
par(mfrow = c(3, 2))
plot(density(unlist(Norm_museum[,2])), main = 'Museum Open Days Distribution')
plot(density(unlist(Norm_museum[,3])), main = 'Museum Operation Hours Distribution')
plot(density(unlist(Norm_museum[,4])), main = 'Museum Rating Distribution')
plot(density(unlist(Norm_museum[,5])), main = 'Museum Total Review Distribution')
plot(density(unlist(Norm_museum[,6])), main = 'Museum Weight Index Distribution')
```











will develop a general function to compute all amenity

will do it later tonight or on Sunday