Data Dinosaurs

Problem Solving Approach

Algorithms used

- We used R, Tableau and Excel for data analysis and visualization

- Combined the tables using R (“sqldf” package)
  - visitation_join <- sqldf("Select * From Member_Visitation Join MedianMeanIncomebyZipCode using (Zip_Code)")
  -
- Pivot tables to group data by months and visualize it through Tableau

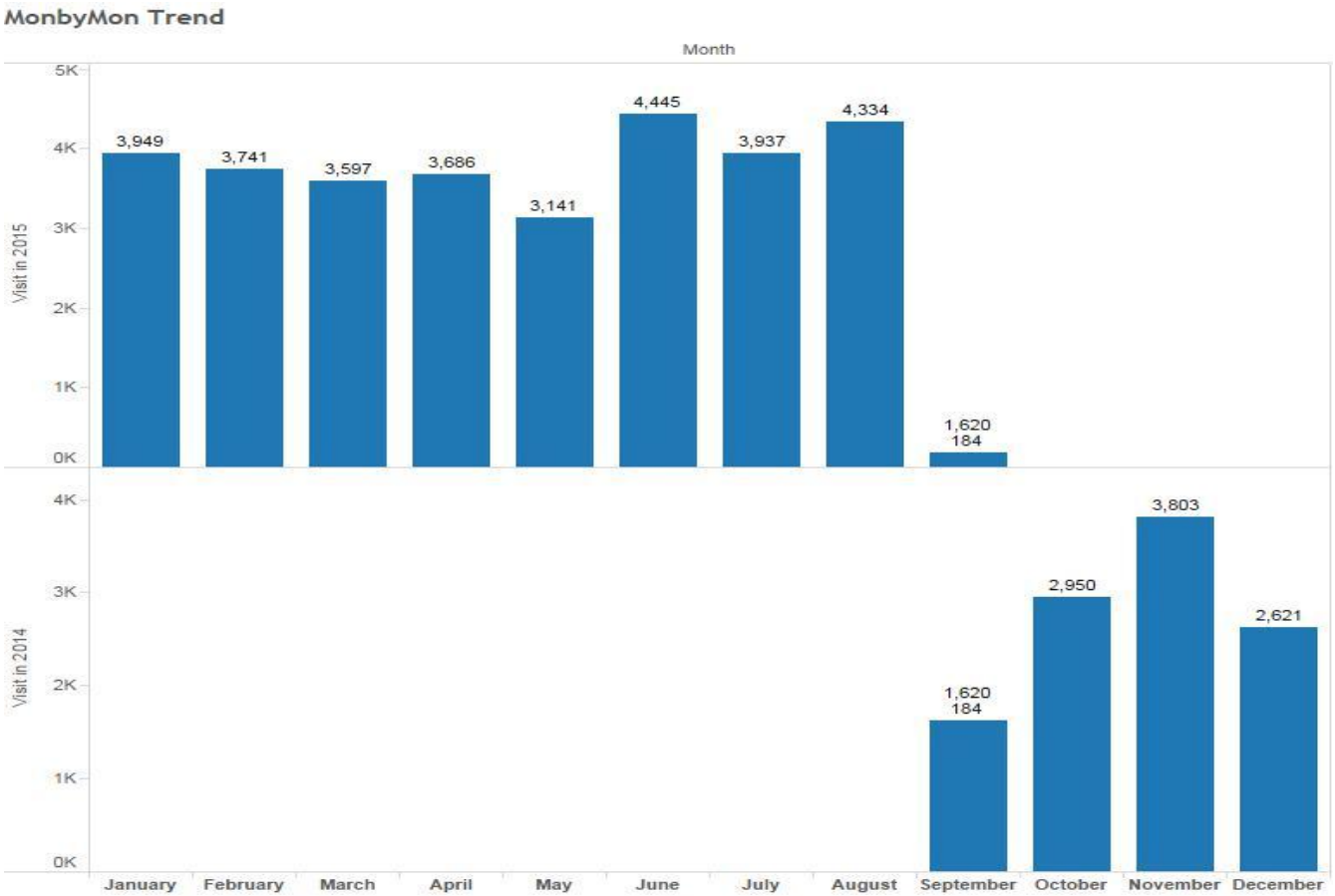
Summary Stats

Summary stats

ID	Level		ZIP_Code	Visitation_Date	Entry_Time
Min. : 1	Family Membership	:37250	Min. : 1002	Min. :41884	Min. :0.210
1st Qu.:2170	Supporting Membership	: 2325	1st Qu.:94121	1st Qu.:42002	1st Qu.:0.420
Median :4247	Family Access Membership:	463	Median :94901	Median :42083	Median :0.450
Mean :4237	Grandparents Membership :	428	Mean :94500	Mean :42083	Mean :0.476
3rd Qu.:6372	Library Membership :	382	3rd Qu.:94941	3rd Qu.:42172	3rd Qu.:0.520
Max. :8434	Inventor :	289	Max. :98112	Max. :42248	Max. :0.950
	(Other) :	871			

Median	Mean	Pop	Month	WEEKDAY
Min. : 23165	Min. : 34293	Min. : 87	Min. : 1.000	Min. :1.000
1st Qu.: 76493	1st Qu.:102378	1st Qu.:11995	1st Qu.: 3.000	1st Qu.:3.000
Median : 85328	Median :125521	Median :29040	Median : 6.000	Median :4.000
Mean : 90698	Mean :128329	Mean :27618	Mean : 6.203	Mean :4.251
3rd Qu.:105815	3rd Qu.:154201	3rd Qu.:38319	3rd Qu.: 9.000	3rd Qu.:6.000
Max. :216905	Max. :336888	Max. :84641	Max. :12.000	Max. :7.000

Month by Month Visitor Analysis



Sum of Visit in 2015 and sum of Visit in 2014 for each Month. The marks are labeled by sum of Visit in 2014 and sum of Visit in 2015. The view is filtered on Month, which keeps 12 of 12 members.

Recommendations

- Which time (month / week) should be recommended to people with children?

  Depends on holidays in schools

- Depending on location, how to recommend?

  Zip codes converted to city names

  The farthest city

  Incentivise them for a open session on a holiday.

Data Cleaning

ZIP CODES

Convert zipcodes table to a table that associate zipcodes with city names using grep and cut and curl

$curl http://ziptasticapi.com/53703

{"country":"US","state":"WI","city":"MADISON"}

zip=”53703”

>command=paste("curl http://ziptasticapi.com/",zip, " ","| cut -d',' -f3 | cut -d':' -f2 | cut -d'\"' -f2 ", sep="")

> command

[1] "curl http://ziptasticapi.com/53703 | cut -d',' -f3 | cut -d':' -f2 | cut -d'\"' -f2 "

> **system(command)**

MADISON

Future ideas

- Use “hclust” R package for hierarchical agglomerative clustering based on the city
- Find the top-k word in the event description and then use the “TwitterR” package to find the top tweets related to event and recommend event to users
- Find the time intervals that the museum has most of the visits and have more open hours in those periods using predictive analytics