
MSIS 670 – MINI PROJECT # 1

TO: Professor Shaffer

FROM: Meghana Aravala

DATE: 24-October-2023

SUBJECT: Analysis on Halloween Visitors Data

I'm excited to share the results of our analysis of the Halloween visitor data, which reveal some fascinating trends in this special dataset. With a focus on making the data understandable to a non-technical audience, this research offers insights into the trends, demographics, and preferences of the visitors throughout Halloween.

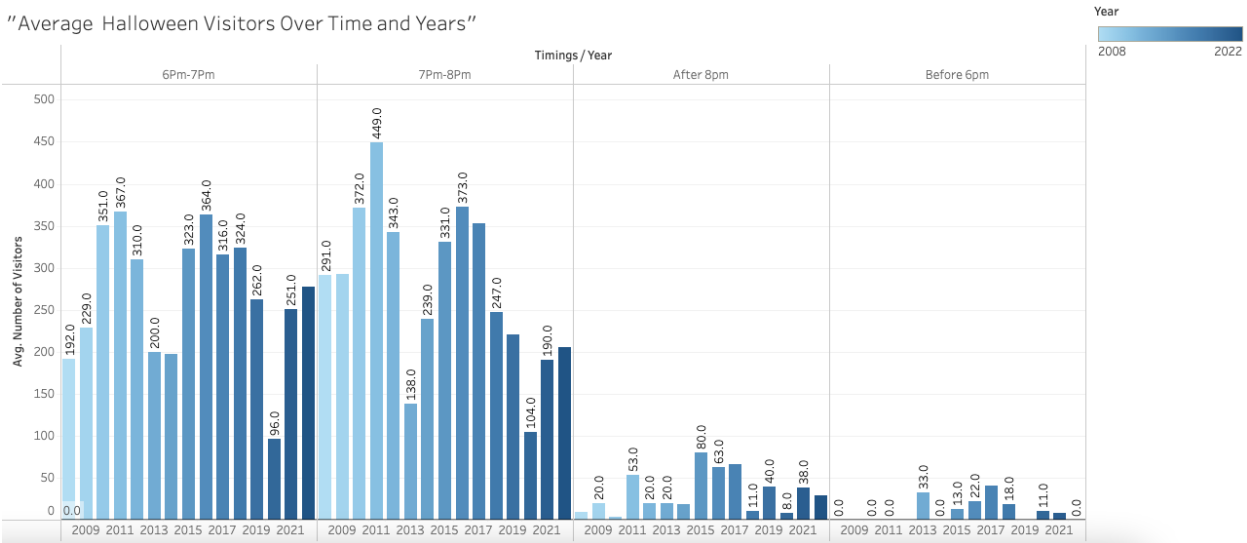
In my exploration of the Halloween visitor data, I used Tableau to generate visuals and derive key insights. The dataset, collected over several years, captures the essence of Halloween celebrations. Below are few of my findings related to various situations.

DATA PREPARATION:

- I started working with a dataset in Tableau that had three Excel sheets named "Halloween 08-15," "Halloween 16-22," and "Day of Week."
- To make things more organized, I first combined the data from the "Halloween 08-15" and "Halloween 16-22" sheets into one dataset using a UNION. Then, I connected the "Day of Week" sheet using RELATION JOIN to add information about which day of the week each Halloween date fell on.
- For a clearer analysis, I needed to tidy up the dataset. So, I pivoted the time-related columns from both "Halloween 08-15" and "Halloween 16-22" into a single column called "Timings." This made it easier to work with and allowed me to see the number of visitors for each time.
- This way, I could get a better understanding of visitor patterns during the Halloween periods from 2008 to 2022

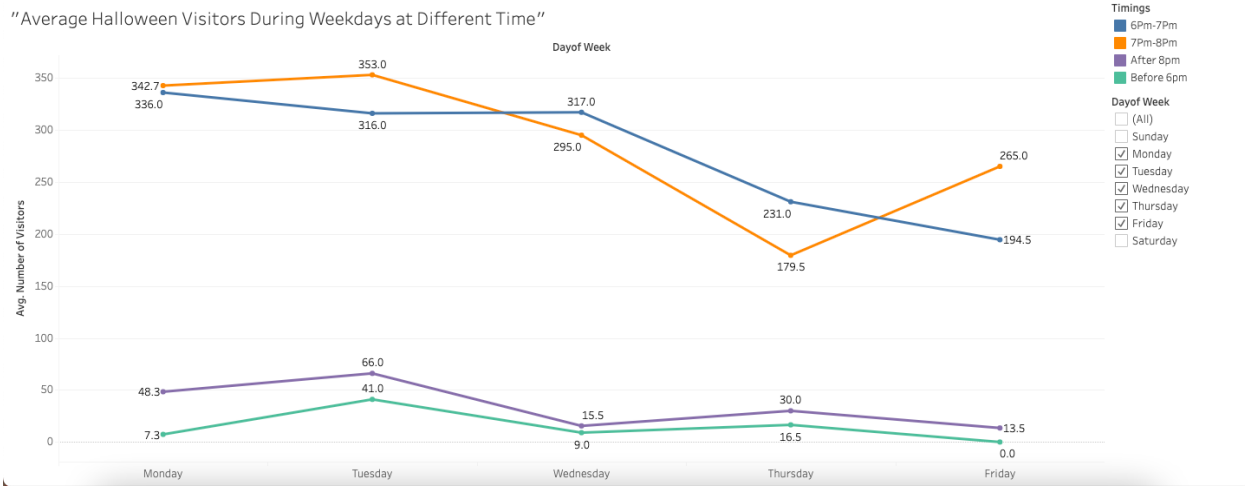
VISUALIZATION 1 (BAR GRAPH) : "Average Halloween Visitors Over Time and Years"

"Average Halloween Visitors Over Time and Years"



- In my bar graph, I've depicted the "Average Halloween Visitors Over Time and Years" by organizing the data in a clear and informative manner.
- The columns of the graph are labeled with "Timings" and "Year," while the rows display the average number of visitors.
- By assigning the "Year" attribute to the color section, I've effectively used color-coding to distinguish between different years.
- By this visualization, we can understand that majority of people would like to visit during 6pm-8pm and very few people are fond of visiting before 6pm and after 8pm.
- We can also observe that the highest number of people visited in the year of 2011 with an average of 449 between 7pm-8pm.
- Least number of people with 0 as an average visited in the years 2008 & 2009 after 8pm.

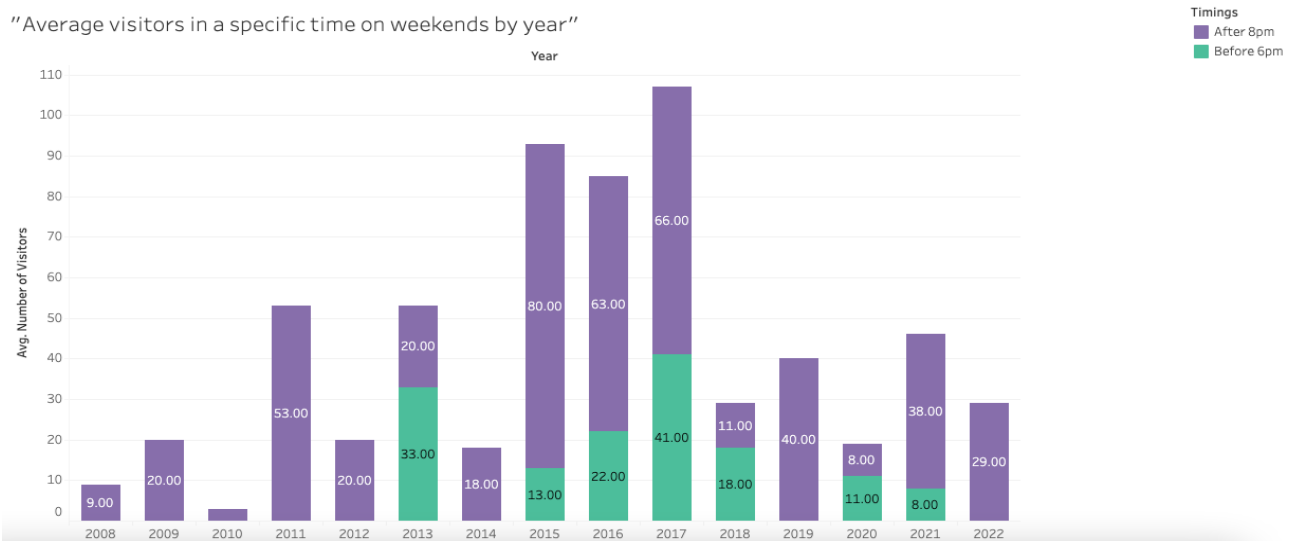
VISUALIZATION 2 (LINE GRAPH) : "Average Halloween Visitors During Weekdays at Different Times.



- In my second visualization, which is a line graph, I've focused on understanding the "Average Halloween Visitors During Weekdays at Different Times." To create this visualization, I utilized the dataset by adding "Day of the Week" as columns and "Average Number of Visitors" as rows.
- To provide a more focused view of the data, I implemented a filter within the visualization. Specifically, I filtered the days of the week and selected only the weekdays which is the main aim of my visualization.
- To enhance clarity and make the data more visually comprehensible, I introduced a color-coding scheme. The "Timings" were assigned as the color attribute, allowing for easy differentiation between different time periods.
- As we observe in 2nd viz, there is slight fluctuation between visitors visiting between 6pm-7pm and 7pm-8pm but when compared between these two, majority of people visited between 7pm-8pm with an average of 353 during weekdays.
- There is a huge difference in between period (6pm-7pm & 7pm-8pm) and (before 6pm and after 8pm). We can evidently say that most of the people do not prefer to visit bef. 6pm and after 8pm during weekdays.
- By comparing 1st and 2nd Viz, we can strongly say that people prefer to visit during 6pm-8pm.

VISUALIZATION 3(STACKED BAR GRAPH): "Average visitors in a specific time on weekends Every year"

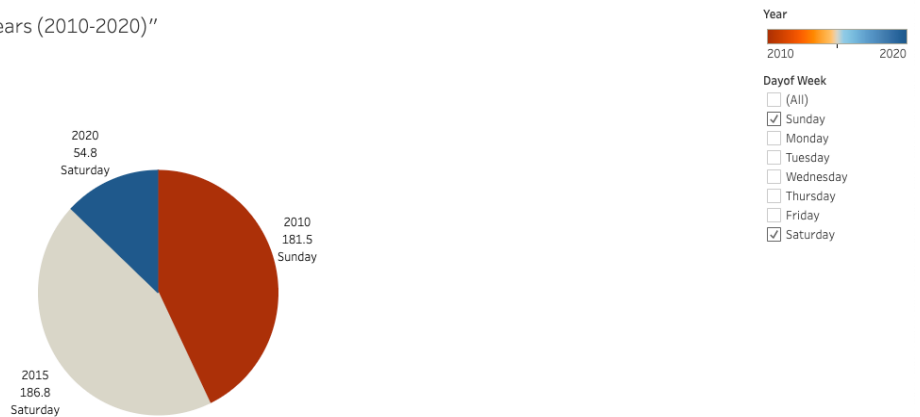
"Average visitors in a specific time on weekends by year"



- In my third visualization, I've employed a stacked bar graph to examine the "Average Visitors at Specific Times on Weekends Every Year."
- To create this representation, I organized the data by placing "Year" in the columns and "Average Number of Visitors" in the rows.
- To provide a more focused analysis, I introduced a filter that allowed me to specify two distinct time periods: "Before 6 pm" and "After 8 pm." This filtering isolates these time frames for a more targeted examination of visitor behavior.
- To enhance the visual clarity of the graph, I leveraged color-coding by adding "Timings" to the color section. This coloring method helps differentiate between the "Before 6 pm" and "After 8 pm" time slots, making it easier to identify and compare the visitor trends during these specific hours.
- For additional context and insight, I included the "Average Number of Visitors" as a text label in the graph. This provides a clear, numerical representation of the average visitor counts.
- We so far saw that the time between 6pm-8pm is most famous visiting hours. I thought of comparing the least busy hours and see which one is least preferred in those 2-time frames.
- We can observe over these years people have not considered to visit the house before 6pm and majority have chosen to visit after 8pm.
- **After 8PM:** The highest average of '80' have visited in this time in the year of 2015 and a least of '3' in the year 2010.
- **Before6pm:** The Highest average of '41' in the year of 2017 and least of '8' in the year 2021.

VISUALIZATION 4 (PIE-CHART): " Average of Visitors during Weekend between Years (2010 -2020)"

" Average of Visitors during Weekend between years (2010-2020)"



- For Visualization 4, I utilized a pie chart to represent the "Average of Visitors during Weekends between Years (2010-2020)."
- To create this visual, I started by selecting the pie chart option from the dropdown menu. I then added the "Number of Visitors" to the angle section, effectively creating a circular chart and changed the measure to Average.
- To provide a more specific and informative view, I implemented filters. Firstly, I included "Day of the Week" in the filters section and chose to display data only for weekends, thereby focusing solely on Saturday and Sunday. Additionally, I added "Year" as an attribute in the filters, restricting the data to the years spanning from 2010 to 2020.
- To enhance the visualization and make it more meaningful, I introduced "Year" as a color attribute. This coloring scheme helps distinguish each year in the pie chart, allowing viewers to easily identify and compare visitor averages for each year within the specified time frame.
- Lastly, for added clarity and context, I included labels within the chart. These labels featured "Year," "Day of the Week," and the "Average Number of Visitors," providing viewers with a clear and comprehensive understanding of how weekend visitor averages have evolved over the years from 2010 to 2020.
- By this pie chart, we can say that from year 2010 to 2020 the event has been held on weekends only in the year 2010, 2015 and 2020 and thus, we get the pie chart in 3 parts.
- Considering only weekends, we can see that majority of people visited on Saturday in the year 2015 with an average of 186.8 , followed by Sunday in the year 2010 (181.5) and Saturday in 2020 (54.8).

VISUALIZATION 5 (CROSS TAB TABLE): Sum of Visitors during Week and Time Every Year

Sum of Visitors during Week and Time Every Year

Year	Sunday				Monday				Tuesday				Wednesday				Thursday				Friday				Saturday				Grand Total
	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	6Pm-..	7Pm-..	After 8pm	Before 6pm	
2008																					192	291	9	0					492
2009																									229	293	20	0	542
2010	351	372	3	0																									726
2011					367	449	53	0																					869
2012													310	343	20	0													673
2013																	200	138	20	33									391
2014																					197	239	18	0					454
2015																									323	331	80	13	747
2016					364	373	63	22																					822
2017									316	353	66	41																	776
2018													324	247	11	18													600
2019																	262	221	40	0									523
2020																									96	104	8	11	219
2021	251	190	38	8																									487
2022					277	206	29	0																					512
Grand Total	602	562	41	8	1,008	1,028	145	22	316	353	66	41	634	590	31	18	462	359	60	33	389	530	27	0	648	728	108	24	8,833

- In my fifth visualization, I created a Cross Tab Table to gain insights into the "Sum of Visitors during Week and Time Every Year."
- To construct this representation, I arranged the data with "Day of the Week" and "Timings" in the columns and "Year" in the rows.
- To quantify the visitor counts accurately, I placed "Number of Visitors" in the label column and adjusted the measure to "Sum." This allowed me to compute the total number of visitors for each year and the specific days of the week and times of day they visited.
- To further enhance the comprehensibility of the table and provide a summary view, I used Tableau's analysis tools. Specifically, I enabled the display of row and column grand totals to see the total number of people who visited each year and, within that, how many visitors came during each day of the week and time of day.
- By using this crosstab table, we can conclude and decide which was the best year, best day, and time so far.
- **YEAR :** We can see majority of people visited in the year 2011 with the total of 869 and least in the year 2020 with total of 219. According to me, the reason for least amount of people visiting in year 2020 can be due to Covid-virus.
- **DAY AND TIME :** Maximum amount of people visited on Monday during 7pm-8pm with total of 1028 visitors and with no major difference comes Monday during 6pm-7pm with the total of 1008 visitors. Least amount of people visited on Friday before 6pm with the Total of 0 and next on Sunday with total of 8 members.

CONCLUSION: According to my analysis, majority of people visited on Monday, Wednesday, Sunday and Saturday and the best timings are from 6pm to 8pm.

<https://prod-useast-b.online.tableau.com/#/site/msis670fall2023/workbooks/973987>

This is the link to my analysis on this dataset in tableau in which I've performed these visualizations.

Thank you,

Meghana Aravala .