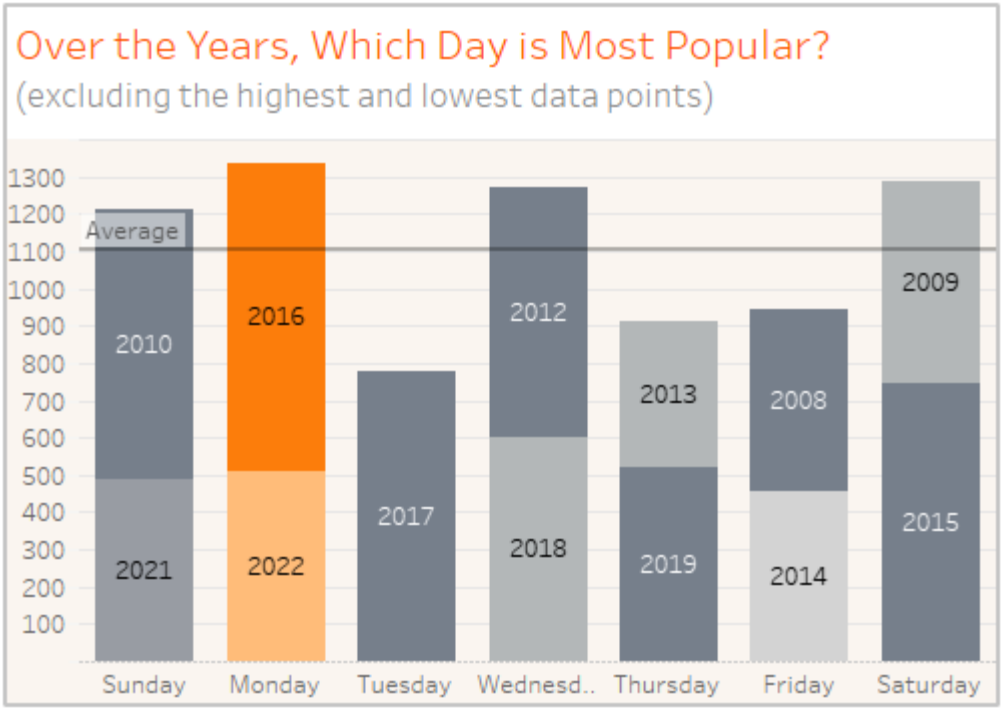
Halloween Visualization

Background: The main storyline setting is in a suburban neighborhood called East Walnut Hills/Evanston which is located in Cincinnati, Ohio. The specific home from which data was obtained is a corner house in the neighborhood with the zip code of 45207 which has an estimated population of 7,974 (from https://www.unitedstateszipcodes.org/45207). There are 2,375 occupied housing units (out of the 3,1888 overall Housing Units) with a Median Household Income of $24,198. According to https://www.weather-us.com/, the Temperature in Cincinnati in October ranged from 64.8 degrees Fahrenheit to 47.3 degrees Fahrenheit with only 10.5 average days reporting rain.

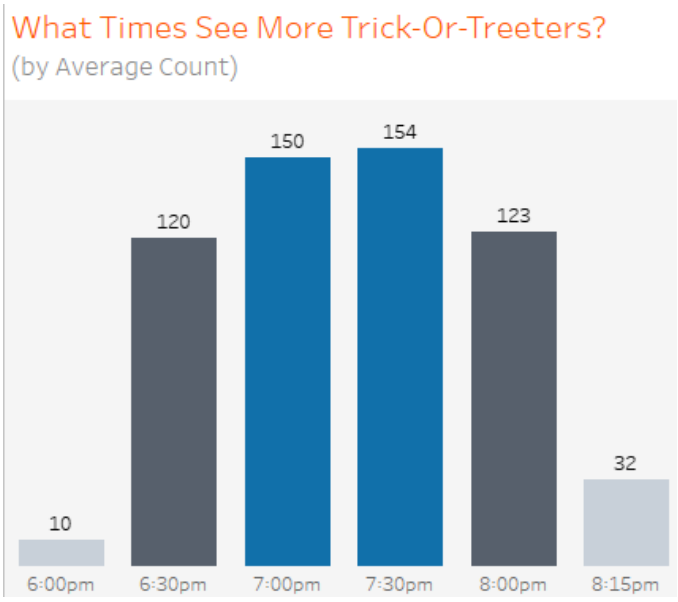
The Storyline basically asks when is the best day and time to expect the most Trick-Or-Treaters and what does the future look like for the next couple of Halloween's?

To answer the first part of the question meant analyzing trends on the Days of the Week, Sunday through Saturday. First, I looked at Frequency charts in both Tableau and Excel using Total Counts of the data collected from 2008 to 2022. This led me to the create a bar chart with Monday's Total standing out among the rest. This led to the exploration of how many days that total stood for. After creating a stacked column chart which demonstrated that there were only two days that had three Halloween dates as input, Monday was the leader by a large margin. So, the stacked column chart was recreated by truncating off the date with the highest count (2011) and the lowest count (2020). That left a chart with 6 days which each had two sets of data and one that only had 1 set of data (from 2008-2022, Halloween has only been on a Tuesday one time). Again, Monday was the busiest date.



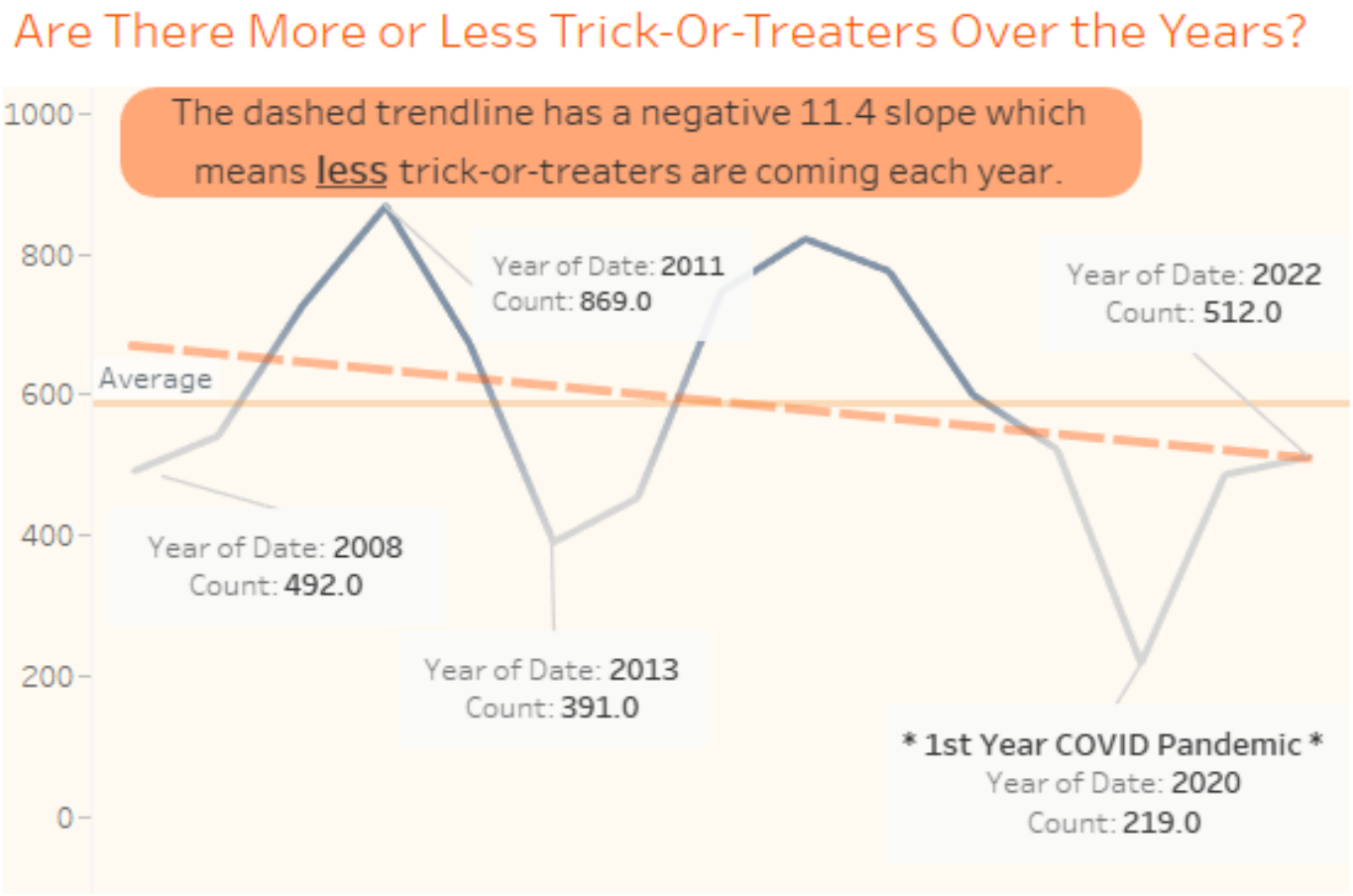
The next item to consider was what time of the day would be the best Time slot to see the most Trick-Or-Treaters. I created a Day & Time Chart that compared the Quantitative data of time slots and the average count of visitors as well as the Categorical Data of the Day of the Week. On the y-axis, I listed the days of the week, on the x-axis the Times slots were listed at the top, and in the Plot Area, there are differently sized Circles to demonstrate the number of visitors. Another preattentive attribute used was cluster groups, signifying that two of the time slots, the 6:00pm and 8:15pm, had the fewest visitors. By using Average Counts instead of actual counts, the fact that our data set had two days with 3 sets of data and one day with only 1 set of data was overcome. This chart was complex but led the

viewer to see that there were 2 somewhat strong indicators in the 7:00pm and 7:30pm time slots that had 4 larger blue circles.



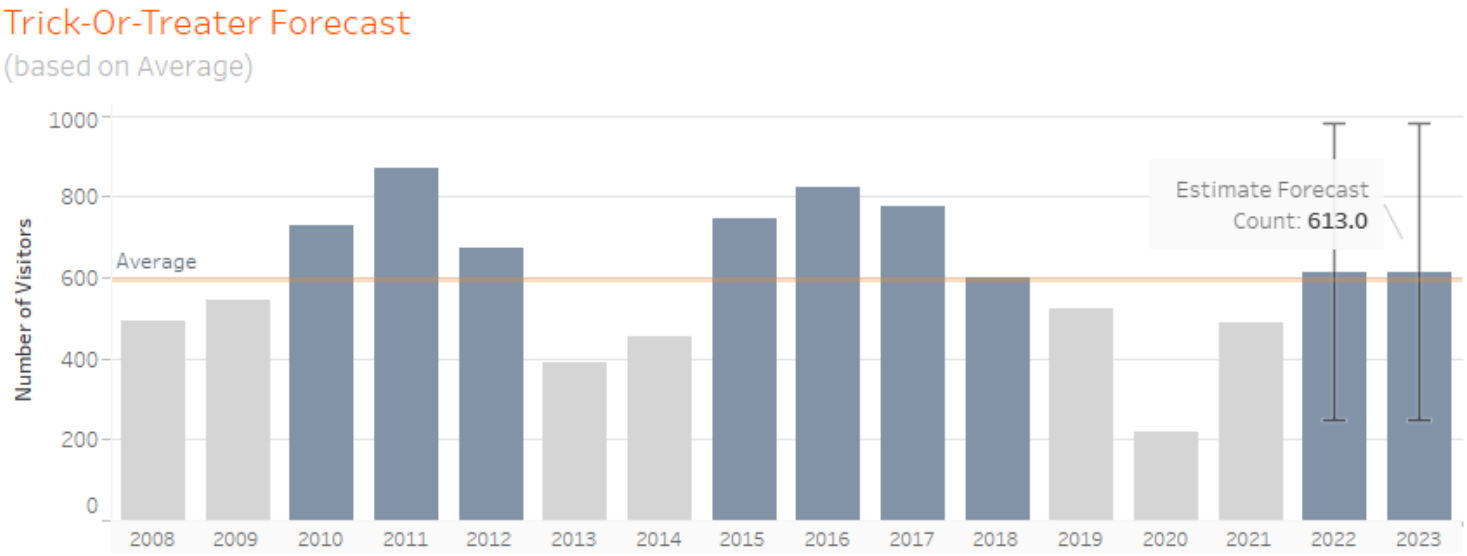
The second Chart called Popular Times was a simple column chart with the Time Slots on the bottom x-axis. The Columns were individually labeled with their Average Count of Visitors. Additionally, they were colored by the same clusters mentioned before, with a brilliant blue standing out as the Peak Times. This chart made it easier to see that the middle cluster(6:30pm and 8:00pm), in a dark grey, were not so close to the high-performing blue cluster. Yet, like before, the low-attendance was noted in the Cluster 3 group which were basically the beginning and ending time for the Trick-Or-Treating event. So, again, the time slots between 7:00pm and 8:00pm were the best times for visitors.

The last main question to answer was what numbers would be in attendance of future Trick-Or-Treating Events in this location. Early on, the Line Graph chart asking, “Are There More or Less Trick-Or-Treaters Over the Years?” was examined to determine the trend in the numbers of visitors over the



years. This chart showed that overall, there is a negative trend in visitors, but it was also noted that COVID had an impact on numbers in 2020. So, a second chart was created, truncating off the lowest and highest values as before. After this change, the trend line was pretty much equal to the Average line.

After being disappointed in the examinations of Running Totals in both Column bar and line graph formats, it was time to go back to looking at the average number of Trick-Or-Treaters per year. This time I chose to keep all the years including the outliers like the low numbers caused by the COVID Pandemic. A simple Column Chart was produced which included the running Average reference line. To help the comparison, the years that were at or above Average were highlighted in blue. The ones below average were dimmed in grey. Additionally, two years of forecasted columns were added representing the next two years based on the running Average. This chart shows that one could expect approximately 613 Trick-Or-Treaters in the future.



In conclusion, this suburban Cincinnati home could see about 613 Trick-Or-Treaters with a majority of them showing up between 7-8pm and they should buy extra candy if Halloween falls on a Monday.

Halloween Visualization

The Assignment

1. Determine a story or goal for the visualization.
   Examples:
   - Homeowner dashboard summarizing Halloween
   - Forecast future trick-or-treaters
   - Explore variation of the number of trick-or-treaters year by year, Peak years, year over year growth, rank, cumulative visitors
   - Time slots - Counts, Cumulative visitors by time, peak time slots by year, variation, changes to time slot counts in two different years (slopegraph, anyone?)
   - Day of week variation
   - Use averages and trends, year over year growth, use color to tell a story, use rank, mark appropriately, use annotations, appropriately. Highlight something about the data.
2. This is a very simple data set. There are only a few years of data broken down into 4 half-hour time blocks with cumulative totals. Think broadly about the data.
   Examples:
   - What comparisons can be made?
   - What table calculations can be made?
   - Be careful about the data - you may need to use averages, but not everywhere.
   - Be careful with the data - **not all days of week are equally represented in the data set** so comparisons across days of week need to be handled so that the comparison for day of week need to be normalized in some way. That is, perhaps there are three events on Sunday and two on Monday. You cannot say that because there were more trick or treaters on Sunday that on average there are more trick or treaters on Sunday because you need to account for the extra day.
3. Build data visualization
   - I generated 13 charts and one heat map for this assignment. Not all of the charts generated are all good. Probably eight charts are good. The heatmap is good. You need to make a lot of charts to find the effective ones. This doesn't even count all the charts that I generated that were immediately rejected for being terrible. I judge my own charts harshly. All of the charts in the final set attempt to use effective color and tell a story.
   - If you create visualizations just to have visualizations that do not add to the analysis or provide a very strong way to understand the story, then you are not doing your job. Creating effective visualizations is a difficult process. The default visualizations will not be good enough.
   - Consider the number of charts generated for the bike share program and previous assignments. There are a lot of charts that try to look at the program from multiple perspectives. Use the kinds of analysis we performed in the past with pivot tables and charts to develop visualizations for this assignment.
   - There are two different kinds of time measurements in Tableau, continuous and discrete. They both make sense in different contexts. Be careful with your dates. Dates, in general, are difficult because we want to do so much with them.
   - Please consider making groups in your dataset to reduce areas that are unimportant - or focus on the important information. Use color effectively to highlight.

Instructions:
Develop charts and tables that tell a story for the Halloween dataset. Use appropriate charts for the given dataset and data. Create one or two dashboards to support your work. Your write up should state how many trick or treaters this address should expect. Provide a range of +/- 50 trick or treaters and your expected value. Your document should support your conclusions (indicate why you chose your range and expected value). Indicate when those trick-or-treaters are expected to arrive (time slot). Complete and submit the assignment by 23:59 EST Sunday.

Halloween Visualization

Brainstorming / Organization

1. **Determine a story or goal for the visualization.**
    - **How many Trick or Treaters should one expect on a given day of the week?**
2. **Think broadly about the data (https://towardsdatascience.com/data-visualizations-with-a-halloween-candy-dataset-b6361d5d29f8)**
    - **Count plots (Categorical Visualizations)**
        - **How many total kids each day of the week? (Stacked bar - Total count on y-axis and 7 days of the week on x-axis)**
        - **Is it increasing or decreasing over the years? (Cumulative Count of kids on y-axis, years on x-axis)**
    - **Box Plots/Violin Plot (Distribution Visualizations)**
        - **Highlight the Mean, Median, InterQuartile Range, Min, & Max.**
    - **Bar/Column charts (Distribution Visualizations)**
        - **Frequency of Data on days of the week (line on graph for each year listed by day of the week)**
        - **Frequency of the Data in specific time slots (% of Total kids on y-axis and time slots on x-axis)**
        - **Rank the Days of the week that will get more kids**
        - **Rank the Day of the week that will get the least amount of kids**
    - **Regression / Scatterplots (Relational Visualizations) – can't do**
        - **Scatterplot/Regression comparing 2 quantitative features between x and y axes (Jittered Scatterplot with years on y-axis and time slots on x-axis)**
    - **Line charts (Clustering Visualizations)**
        - **What is the trend over the years: more kids or less kids?**
    - **Future data exploration thoughts**
        - **How are the rise of food allergies impacting trick or treaters?**
        - **Is the fear of COVID/Other contagious disease slowing the trend?**
        - **Which kind of candy is preferred?**
        - **What kind of candy is NOT preferred?**
        - **How has inflation impacted the number of homes giving out candy?**
        - **When is the candy for Halloween usually bought?**
        - **What ages are you seeing trick or treaters?**
        - **What percentage are dressed up in costumes?**
3. **Build data visualization (see page 1-3).**
4. **Write up how many trick-or-treaters this address should expect. See page (1-3).**