Trends in Seattle Library Usage

Taylore Prinsen
taylore.prinsen@colorado.edu
Colorado University at Boulder
Boulder, Colorado

Sky Johnson
sky.johnson@colorado.edu
Colorado University at Boulder
Boulder, Colorado

Peter Minwegen
peter.minwegen@colorado.edu
Colorado University at Boulder
Boulder, Colorado

Figure 1: The Seattle Public Library

ACM Reference format:
Taylore Prinsen, Sky Johnson, and Peter Minwegen. 2019. Trends in Seattle Library Usage. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 10 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 ABSTRACT

This study was an attempt to extend research on checkout trends through the year of the publicly available Seattle Library system. We sought to find out if there was any noticeable correlation between certain weather patterns, and library usage.

We looked to see if the rate at which assets within the library were checked out at were influenced by average temperature, rain and precipitation, wind, and visibility. We found that there was no noticeable change or correlation between these and asset usage.

2 INTRODUCTION AND BACKGROUND

The Seattle Public Library is a landmark of the city for it's architectural style, and unique community gatherings. However we wanted to look further into it's functions as an actual library.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2.1 Problem Statement/Motivation

The Seattle Public Library is a landmark of the city for it's architectural style, and unique community gatherings. However we wanted to look further into it's functions as an actual library. Before embarking on this project, it was unclear whether or not trends in rentals of library books (or other materials) in Seattle are impacted by the weather trends in the area. Our goal was to look at data sets both of library checkouts and weather in the Seattle area and look for correlations and trends. We wanted to explore whether or not people check out books more or less when the weather is rainy, sunny, etc. Our goal of this project was to better understand the implications that weather can have on the practices and habits of library users.

This question is important in the justification of building public facilities like these throughout a city. Often times a city council will have to determine if a city will benefit from spending it's budget on projects like these, or other infrastructure. In the past it used to be a more clear decision, since they would all see usage. However, looking at our initial studies on library checkout usage will show a general trend that the library use is in decline. Hopefully determining if erratic weather patterns, or temperate climates play a role in how a city will use it's public facilities will be beneficial when making such investment decisions.

Further, many people suffer from Seasonal Affective Disorder (SAD) and are left facing varying side effects from weather impacting their mood and demeanor. We are intrigued by the ways that weather can impact people's moods and behaviors and wanted to begin to understand how this may impact people's daily habits, going to the library, etc. This could bring great insight to the behavior of people facing seasonal depression.

https://www.mayoclinic.org/diseases-conditions/seasonal-affective-disorder/symptoms-causes/syc-20364651

3 LITERATURE SURVEY

There has been very little research done with these data sets in the past. The only work we came across that used the Seattle Public Library data set only referenced data specific to that database - what time of day is most popular for rentals at the library, which type of rental is most popular at certain times, etc. At the end of this quick study the author came to the conclusion that 4pm, Fridays in January and July are the busiest days for activity within the library's checkout system data set. However this is limited and doesn't ask much of why that is the case. The intent of our project was to expand on these findings with correlation to various weather data. This work was done by Yorgos Askalidis in January of 2018. Below are some of his graphs that he provided on his article.

Figure 2 (shown below) shows the percentages of the checkouts in each day per hour - you can see that the most popular hour for checkouts is around 4 p.m. on average. The slowest hours of the day include 10 a.m., 6 p.m. and 7 p.m.. These findings show that in the beginning and ending of the days is when it is the slowest - this would make sense as people are typically most active around the peaks of the day. The high activity around 4 p.m. also makes sense since that is when people are coming home from work - most likely time for them to stop at the library. This is the most relevant information for our project - as we focus mainly on the hourly data in our bulk of the work and less on other time factors.

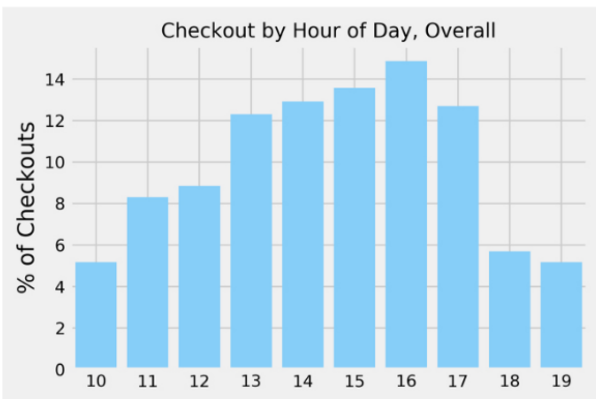

Figure 2: Number of Checkouts (Hour of Day

Askalidis also did some analysis on the day of the week and the month of the year regarding library checkout activity. Figure 3 shows us that most of the days are fairly equal in the amounts of checkouts the library has. Monday through Thursday are the most constant days, with a small dip in the amount of checkouts on Friday, followed by an increase in checkouts on Saturday and another dip on Sunday (the least amount of checkouts of all the days of the week).

Figure 4 shows us the checkouts by months in the year. It seems that the checkouts stay almost consistent across the months of the year - which can clue us in that temperature may not have a huge effect on the amounts of checkouts we can expect.
These findings can help us understand the trends of the checkouts without the presence of weather data. We can understand at a very broad level how weather could interact with library checkouts. Knowing that in all the months of the year the checkouts remain constant, we can infer that the cold temperatures may not matter - but this leaves the question of rain (precipitation), etc. has anything to do with weather trends by the hour or day of the week.
Some of the other trends that Askalidis found with the Seattle Library Data include:

- Number of checkouts across the years (separated by books and DVD/CD)

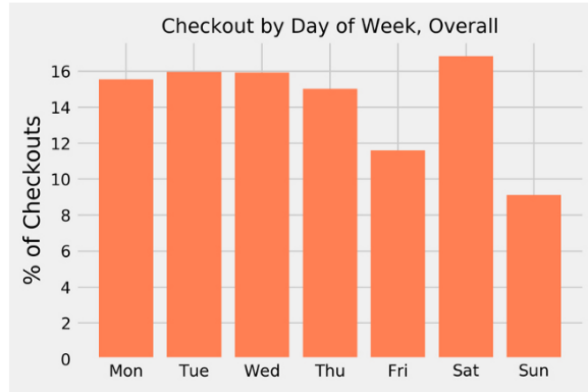

Figure 3: Number of Checkouts (Day of Week)

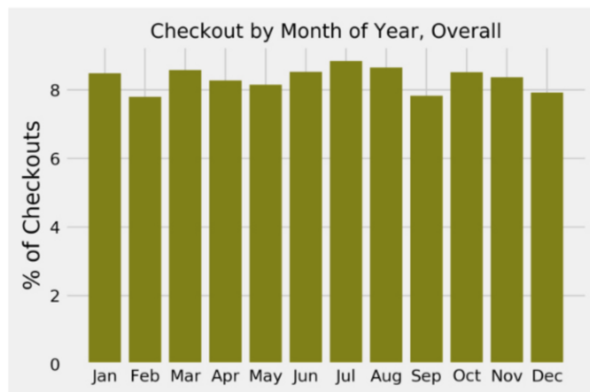

Figure 4: Number of Checkouts (Month of Year)

A steady number of book checkouts (around 4 million) observed, while CD/DVD checkouts decreased over time, peaking around 2009 - probably from the emergence of Netflix, Spotify, etc.

- Checkout temporal trends:
  - Time of Day
    4 p.m. is the most popular hour for checkouts - possibly because people get off work and then head to the library?
  - Day of the Week
    Saturday has remained the most popular day for all years in the data set (Friday and Sunday are the least popular days).
  - Month of the Year
    "The beginning of the year and the summer seem to be the most popular months for checkouts, with September and December the least popular." (CITE - from toward data science link)

https://towardsdatascience.com/how-and-when-people-use-the-public-library-1b102f58fd8a

4 DATA SETS

We used two data sets for our main research. We used one data set that shows us the checkout trends at the Seattle Library and another data set that showed the weather by day and hour in the Seattle area

4.1 Seattle Checkouts by Title

This data set contains information for the Seattle Public Library for physical and digital data. There are 1,431,563 unique objects in the data set, each with data for 11 attributes:

- UsageClass (Nominal)
  Denotes if item is âĂIJphysicalâĂİ or âĂIJdigitalâĂİ
- CheckoutType (Nominal)
  Denotes the vendor tool used to check out the item.
- MaterialType (Nominal)
  Describes the type of item checked out (examples: book, song movie, music, magazine)
- CheckoutYear (Numeric/Interval)
  The 4-digit year of checkout for this record.
- CheckoutMonth (Nominal)
  The month of checkout for this record.
- Checkouts (Numeric/Interval)
  A count of the number of times the title was checked out within the âĂIJCheckout MonthâĂİ.
- Title (Nominal)
  The full title and subtitle of an individual item
- Creator (Nominal)
  The author or entity responsible for authoring the item.
- Subjects (Nominal)
  The subject of the item as it appears in the catalog.
- Publisher (Nominal)
  The publisher of the title.
- PublicationYear (Numeric/Interval)
  The year from the catalog record in which the item was published, printed, or copyrighted.

This data set begins in 2005. Link: https://www.kaggle.com/city-of-seattle/seattle-checkouts-by-title

4.2 Seattle Weather Data

4.2.1 *Did It Rain in Seattle? (1948-2017).*

"Besides coffee, grunge and technology companies, one of the things that Seattle is most famous for is how often it rains. This dataset contains complete records of daily rainfall patterns from January 1st, 1948 to December 12, 2017."
This data was collected at Seattle Tacoma International Airport, WA.
https://www.kaggle.com/rtatman/did-it-rain-in-seattle-19482017
https://www.wunderground.com/hourly/us/wa/seattle/KSEA
https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND

4.2.2 *Seattle Weather csv.*

This data set contains 1463 unique objects. Each is a day included in the range from January 1 2012 to December 31 2015. This data was found on Github:
The attributes for the objects include:

- Date (Numeric/Continuous)
  The unique day, month, year for each entry
- Precipitation (Numeric/Ratio)
  Amount of precipitation.
- TempMax (Numeric/Ratio)
  The maximum temperature that day.
- TempMin (Numeric/Ratio)
  The minimum temperature that day.
- Wind (Numeric/Ratio)
  The wind speed in mph.
- Weather (Nominal)
  Drizzle, Rain, Sun, or Snow

This data was found on Github: https://github.com/domoritz/maps/blob/master/dat weather.csv

4.3 NOAA Local Climatology Data for Seattle Tacoma Int'l Weather Station

The attributes for the objects that we included in our pruned data set include:

- Date (Numeric/Continuous)
  The date and hour for each entry
- HourlyAltimerSetting (Numeric)
- HourlyDewPointTemperature (Numeric)
  Dew point temperature at each hour
- HourlyDryBulbTemperature (Numeric)
  Dry bulb temperature at each hour
- HourlyPrecipitation (Numeric)
  Precipitation at each hour
- HourlyPresentWeatherType (Numeric)
  Weather type at each hour
- HourlyPressureChange (Numberic/Percent)
  Pressure change from this hour compared to last
- HourlyPressureTendency (Numeric)
  Tendency of pressure hourly

https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:24233/detail

5 TECHNIQUES AND METHODS APPLIED

Due to the exploratory nature of this project, we had to revise and plan to explore more options of analysis as we continued to work on this.

5.1 Data cleaning

5.1.1 *Data Pre-Processing.*

To manage the data and pre-process it. We will mostly be using Panda's data frames and it's convenient features. Specifically, we will use it to quickly import, load, join, and edit columns of various data sets, and change various indexing attributes. This will allow us to spend minimal time editing and pre-processing the data. And allow us to drive straight into the project sooner.

5.1.2 *Time Series Analysis.*

The next big challenge that we had to approach is how we handled time series within the data set. Again, we used Data Frame objects provided by Pandas to separate, join, and combine specific portions of data . For example, we will need to get rid of hours the Seattle Public Library isn't open to the public, since there is little likelihood that these will play any part in determining when a book gets returned. We would solve for something like this, by setting the Checkout time to the Panda's frame Index. And by selecting only 8am to 6pm, we can be assured we have covered all open hours of the library system.

5.2 Evaluation Methods

How we evaluate and approach the data will obviously change as the project continues.

We will start however, with the aim to look into discovering interesting correlations between data based on time series analysis techniques. One of the biggest challenges we will face is trying to differentiate between when there is a significant event that has a correlation to another pattern, vs random noise.

I think it will be appropriate to measure how successful our project is based on how many interesting correlations we can find, and how we can support them with statistical significance. However we won't be able to figure out this amount, or if it's possible at all, without more work on the data set.

5.3 Tools

For this project we will use the standard libraries with Python and Jupyter Notebooks. Specifically, we plan to import and handle our data using Panda's DataFrames so we can quickly sort through and evaluate large portions of the data at a time.

We can then use Numpy's tools and methods to create any confusion matrices and arrays, while also letting us handle the large data sets without too much trouble.

Finally for Time Series, we can make use of the stats models API library to quickly get started in looking for statistically interesting events within the time series data, and finally we will use MatPlotLib to visualize.

6 KEY RESULTS AND ANALYSIS

6.1 Initial Data set results and exploration

Our first goal and area of work was to understand and visualize the Seattle Library Checkouts Data Set individually of the weather data. By doing so, we could begin to understand the typical trends and patterns present in the library checkouts of Seattle. And if we would need to rethink any of our approaches and datasets. We started by diving right in, and trying to find any pattern between rain and checkout numbers.

When analyzing checkout numbers, we began by following prior work here: https://towardsdatascience.com/how-and-when-people-use-the-public-library-1b102f58fd8a . We soon identified number of checkouts, checkout object type (Book, DVD, CD, VHS, Other), checkout time by day, checkout time by month, and checkout day of the week as points of interest to focus on. From each day, week, and month analysis; there will be 3 graphs. graph overall, graph by object type, and graph over all years. These 3 types of graph for each day, week, and month should provide excellent analysis over the checkout data. Another possible avenue to explore could be titles of each object type to determine a category (such as Fiction, Nonfiction, Mystery, Kids, Etc...). We could cross examine the category by time of year to see if peoples interest in Books, DVD's, CD's, or VHS change depending on the time of year (or even week).

6.1.1 Quick Analysis of Rain + Checkouts. We wanted to get a quick and brief look in to whether or not weather (more specifically, rain) drastically affected the number of checkouts in the Seattle library. Since the Seattle library checkouts data has millions of data points, to cut back on run time, we looked just at the years of 2016 and 2017, as they seem to have just around the same amount of checkouts each year (right around 6 million). We also used the "Did it Rain in Seattle?" data set that has data on whether or not there was rain in Seattle each day between the years 1948 to 2017. We pruned both the Seattle Library Checkouts as well as the "Did it Rain in Seattle?" data to include just the years 2016 and 2017. At first, we were looking to include 2011 to 2017, however, once calculating the run time for certain functions, we decided to look at a smaller data set. We understand that this is a very small amount of data and will provide no conclusive results, however, we still wanted to get a glimpse into whether or not the two interact at first glance.

The Seattle Library checkouts has two data sets provided - one that looks at both physical and digital checkouts (the one we used to get a quick glimpse) and one that looks at only physical checkouts (which we will use for further exploration). One caveat of the digital/physical data set is that the data for each checkout only includes a month and year for checkout, rather than a day, month, and year. Because of this, we decided to create a function to determine whether each month (January - December) in both 2017 and 2016 was **on average** rainy or not. To do so, we created a nested for loop that would look through the amount of rainy days that we counted in each month (in each year) and returned a boolean True/False value indicating whether or not each month was **on average** rainy or not.

Once we had data regarding whether or not each month was on average more or less rainy, we appended the boolean value on to the original data set in a new attribute labeled IsRainyMonth and created a new csv file ('checkoutWithRainy.csv') to be used to further look at the correlation between the data. We used various pandas functions to look at the data in various ways visually. We plan to work a little more on this as time permits, however, we have found that there seems to be no initial correlation between rain and library checkouts in this small data set. The graph below shows the months January through February (labeled 1-12) as well as the two years of 2016 and 2017. You can see that the only spike in checkouts with the presence of rain was during the months of January and February. The following graph shows the months January through December (labeled 1-12) as well as the years 2016 and 2017 (labeled 16 and 17). The only slight trend that we could pick out was the possible correlation with rain and checkouts in January and February in both years - though the results prove inconclusive at the current state.

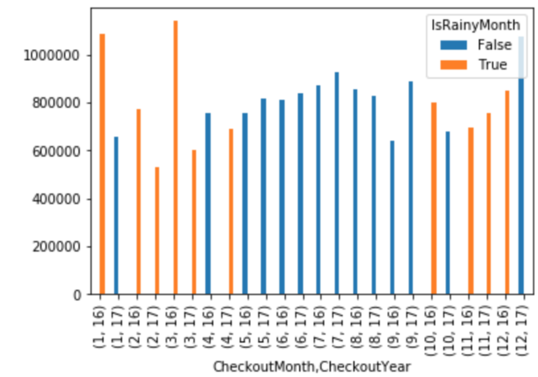

Figure 5: Rain and Checkout numbers

6.2 In depth looks with better data sets

After our initial exploration we realized some things about our data sets, and had decided to switch our approach slightly, we realized some limits with the initial Data set provided by kaggle for library checkouts.

6.2.1 In depth look at Physical Checkout trends of Seattle Public Library . We found another data set that included hourly checkouts since 2005, so to gain a better understanding of the library data we decided we would split and partially focus on dealing with that data (See below). We also found that this data we could pull directly from the Seattle Government site, and as a result we would have much more up to date data. The new data is located here : https://data.seattle.gov/dataset/Checkouts-by-Title-Physical-Items-/3h5r-qv5w

Another data set utilized for looking at all the physical item checkouts from Seattle Public Library was found here: seattle-public-library-checkouts-kaggle" This data set begins with checkouts occurring in April of 2005 throughout September of 2017.

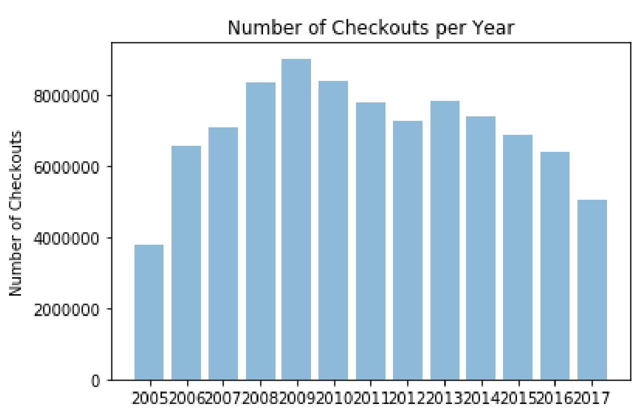

Figure 6: Number of Checkouts

In the graph above, it is important to note that this data set began in April of 2005 and ended in September of 2017. This explains the vastly reduced number of checkouts apparent in the years 2005 and 2017. We can see that the largest number of checkouts occurred in 2009. The number of checkouts is the total count of checkouts for any object type (Book, DVD, CD, VHS). We notice an increasing number of checkouts from 2005-2009. Afterwards, number of checkouts steadily declined yearly (except 2013; can possibly analyze special weather circumstances for the year 2013) Possibly a growing use of technology is responsible for the recent declines in the Seattle library checkouts. It should also be noted that the Seattle library recording this data has different hours based on different days of the week. Friday and Saturday have slightly reduced hours of operation while Sunday has greatly reduced hours of operation.

Data set graphs on time of day in creation...

When looking at the time of day, it is crucial to understand that the library opens at 10am but closes at 8pm. It closes at 6pm on Friday and Saturday, while on Sunday it closes in range of (1pm-5pm). However, these library hours are only the CURRENT library hours. They may or may not have been different in the past (the graph could potentially point to this)

graph:Checkouts by hour overall

graph:Checkouts by hour by object type

graph:Checkouts by hour across all years

Data set graphs on time of week in creation... The library is open every single day of the week

graph:Checkouts by week overall

graph:Checkouts by week by object type

graph:Checkouts by week across all years

Data set graphs on time of month in creation...

It is important to understand the potential impact of the months containing a varying number of days (28-31). Each month also contains a different number of each day of the week, this could impact the total number of open hours for the library.

graph:Checkouts by month overall

graph:Checkouts by month by object type

graph:Checkouts by month across all years

6.3 In depth look at weather trends of Seattle

We also came to a similar conclusion involving weather data, we found it extremely hard to find accurate data from Wunderground since they no longer have an API, and web crawlers were taking too long, so we moved our focus to use NOAA's Local Climate Station data sets, specifically, the data set from Tacoma Airport, just south of the majority of the Seattle Public Libraries. This data set has 100% coverage of temperature and other various data points since 1944! (Although, we only used from 2005, the date the library data set goes back too). You can also find the data set here : https://www.ncdc.noaa.gov/cdo-web/datatools/lcd

Before we combine the results of the Library analysis, with weather data, we decided it would be helpful to understand some basic trends of the climate at Seattle. And if there has been any noticeable changes within the last 5 years.

First, we will look at how the temperature averages changed over the last few years to see if there is anything to be discovered there.

Below is a graph of the Monthly average temperatures for the last few years. Made by re-sampling the "HourlyDryBulbTemperature" to monthly using Pandas.

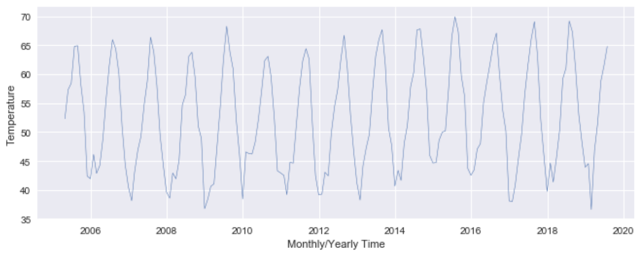

Figure 7: Average Monthly Temperature

As we can see from the graph (Figure 3) there appears to be a fairly expected trend, where December will report the coldest months, and July/June the hottest months. We can compare these two seasons to see if there is a noticeable change in checkouts in the next portion of this study.

Next, we can look at the rainfall/precipitation in more depth since 2005. To do this we will need to look at the hourly precipitation column. Originally values of "T" will indicate trace amounts, so we just replaced these values with 0.

Anyway, re-creating a similar graph to the temperature yields less intuitive results!

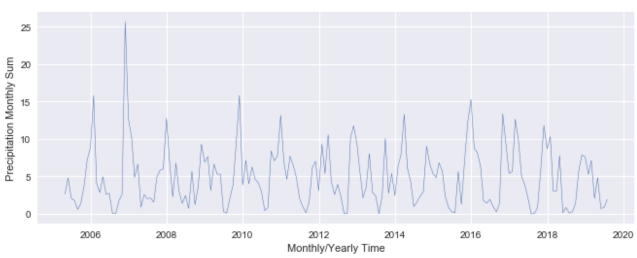

Figure 8: Monthly Sum of Precipitation

Observation of the graph shows that the months and amounts of rain vary a whole lot more than the temperature, which will lead to some more interesting analysis when we combine them with checkout numbers.

To get a better idea of the "rainy" Months, we broke it down into weeks, and plotted all the years against each other.

This graph is a little bit hard to interpret but we can see that the months with the least amount of precipitation are mostly during the summer, while November and January tend to receive substantially more. It is important to note that precipitation includes rain AND snow in this case.

We will have to include the different types of weather in future

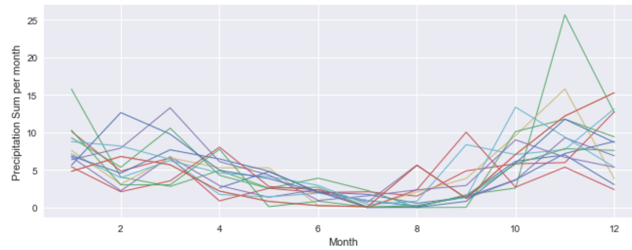

Figure 9: Month Sum of Precipitation

```
In [32]: display(WeatherData_R['HourlyDryBulbTemperature'].corr(CheckoutData['ItemType']))
         display(WeatherData_R['HourlyPrecipitation'].corr(CheckoutData['ItemType']))
         display(WeatherData_R['HourlyVisibility'].corr(CheckoutData['ItemType']))
         display(WeatherData_R['HourlyWindSpeed'].corr(CheckoutData['ItemType']))

         0.10614282121071186

         0.002912247874910502

         0.07339364882775207

         0.06130736462640944
```

Figure 10: Notebooks correlation results

analysis for this project since it is likely to play an important role in determining attendance to Libraries.

6.4 Correlation between weather and checkout patterns

At this point in our study we decided to combine the data sets from checkout numbers and weather data. The first step was to reduce the total size of the library checkout data since it was entirely to large for practical usage. And since most the other information and columns besides date were irrelevant to whether or not there was an influence of weather.

To start this, we again, broke up the checkout data into smaller more manageable parts using the Unix command "split", this way we wouldn't have to load the full 26 gigabyte CSV for checkout data at the same time. We then re-sampled the checkout data for every hour to match with the hourly weather data. As a result this also vastly decreased the overall size of the data set.

However, we had to make trade-offs on what information we kept. We decided to just prune most of the other features so we were just left with the number of checkouts per hour for every hour.
pruning excess data, we then pruned the time-frames using Panda's date interpretation indexing. This allows us to look at the data sets between the time ranges from 8am - 6pm.

We then merged this re-sampled data frame with the NOAA data set, and started looking at the raw correlation values. These were our results.

As we can see with these results, the strength of the correlations between the amount of checkouts per hour, and that hour's current weather state don't seem to play that large of a role.

We decided however, that tbhis could cause some error, or over simplification of the natural library usage trend through a day however. This correlation result looks at weather and checkout

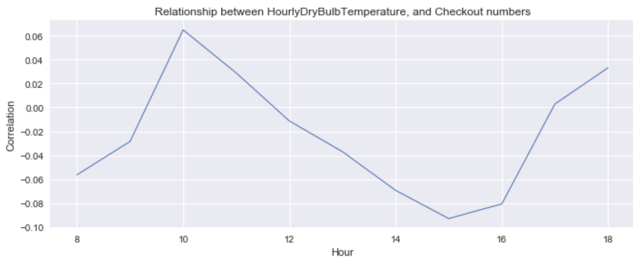

Figure 11: Correlation results through the day for Dry Bulb Temp

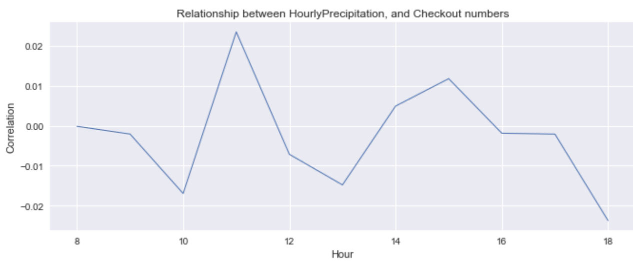

Figure 12: Correlation results through the day for Hourly Precipitation

numbers for all hours between 8-6pm indiscriminately. However, we already know that the least used hours of the library are during the opening hours at 8am. We decided that we could break up the day more individually.

To do this, we split the weather and checkout data set on every hour. Making a smaller data frame for 8:00 to 8:59, 9:00-9:59...etc. This way, if there was a correlation between a weather feature, it would be more representative of standing out depending on that exact hour. However when we did this we discovered our correlation to weaken even more.

As we can see, the highest correlation results we get in this case are barely above 5% which shades more doubt on the case that the checkouts and weather share a clear correlation.

6.4.1 Principle Component Analysis. As another measure to see if there was some value with our data set, we also wanted to quickly run Principle Component Analysis to see if there was a single feature that stands out more than the rest.

PCA works by decomposing the Eigenvectors of a given matrix, to determine the most important "directions" or columns in said matrix. We can apply this to the weather data.

To do this, we used Python’s SciKitLearn PCA object. We took the weather data and created a numpy matrix. Then we used sklearn’s StandardScaler object to refit the data for PCA. We have to standardize the dataset otherwise naturally high columns will overtake the principle components.

After applying the principle component analysis, we got the following results for the weights of each column.

As we can see, the biggest value is almost responsible for 20% of the feature representation. But, about 80 of the data can be

```
pca.explained_variance_ratio_

array([2.09789149e-01, 1.22835798e-01, 1.08404412e-01, 9.22962764e-02,
       8.82804738e-02, 6.23760083e-02, 6.15923842e-02, 5.23672557e-02,
       4.73106943e-02, 3.97412014e-02, 3.51492941e-02, 3.03322985e-02,
       2.58277148e-02, 2.26050997e-02, 5.84798957e-04, 5.07141001e-04,
       1.13788534e-35, 0.00000000e+00])

pca.singular_values_

array([4.38743444e+02, 3.35723395e+02, 3.15386143e+02, 2.91012216e+02,
       2.84610853e+02, 2.39236764e+02, 2.37729259e+02, 2.19204284e+02,
       2.08352536e+02, 1.90958777e+02, 1.79588049e+02, 1.66829094e+02,
       1.53943771e+02, 1.44019846e+02, 2.31644730e+01, 2.15716563e+01,
       3.23123378e-15, 0.00000000e+00])
```

Figure 13: Notebooks values for Principle Component Analysis on Weather data

represented with about 8 features, (out of 18). That naturally means some features are way more deterministic for representing the data than others. However this doesn’t necessarily play a part in determining whether they effect library checkout information on their own.

7 APPLICATION

Since the results of our research ended up showing no significant correlation between the weather and checkout patterns in Seattle, there isn’t really too much to apply our findings to. Since it shows that Weather data doesn’t impact immediate library checkout results in the Seattle area, it could prove to be beneficial for City Council to look at these results in order to decide whether or not it could be beneficial for them to invest in public facilities.

Although we did not garner the results we were anticipating, actionable insights were still made. In our research we came to realize that checkouts peeked in 2009. While this has been mentioned previously, one facet that has been over looked is that of the a potential cause for the decline afterwards. We believe potential driving forces for this decline include streaming services such as: Netflix, Amazon, and Hulu. We also believe that the creation and mass adoption of digital devices such as smart-phones and tablets aided in the decline of checkouts. In 2006, Amazon offers video on demand as a service. In 2007, Netflix announced that it will launch streaming video. Also in 2007, the Apple iPhone and Amazon Kindle were released. In 2008, Hulu launches for public access in the United States. We believe it took a small period of time spanning roughly 3 years for the American people to become accustomed to the these new digital streaming services and devices.

The graph below was developed by The Seattle Times with data sourced from the US Census Bureau.

population chart.jpg

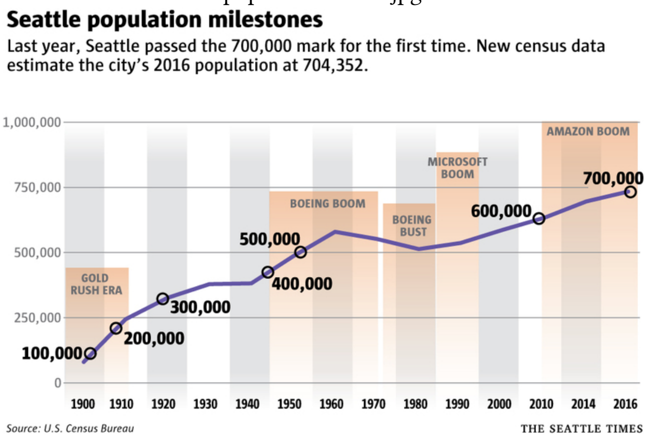

Figure 14: Seattle Popluation

From the graph above, we can see that the population has been increasing rather steadily since the Boeing bust in 1980. This population growth was supported by industry giants Microsoft and Amazon headquartering in the area. This population growth is important to note because it reveals that population plays no part in the decrease of the number of checkouts per year. This confirmation is in line with our expectation that the decline in the number of checkouts is due to streaming services.

Furthermore, lets take look at the population of various digital devices from the year 2004 to the year 2015. The graphs below are courtesy of Pew Research. The graphs represent the percentage of American adults who own the digital device.

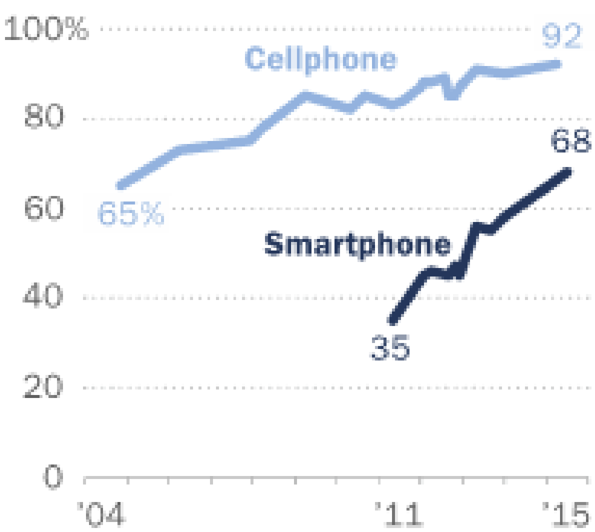

Figure 15: Percentage of American adults who own smartphones

When we take a moment to look at this chart it is apparent that smartphone ownership doubled in 4 years and this demonstrates the rapid adoption of digital devices. It is also important to note that this massive smartphone adoption began in 2011. Although, 2011 is the earliest year that Pew Research began collecting the data. There may or may not have been an increasing trend in smartphone usage in the years 2009 (year of maximum library checkouts) and 2010. However, it would appear that cellphone ownership in general increased in the years 2009 and 2010. Thus it is likely that smartphone ownership also trended upwards.

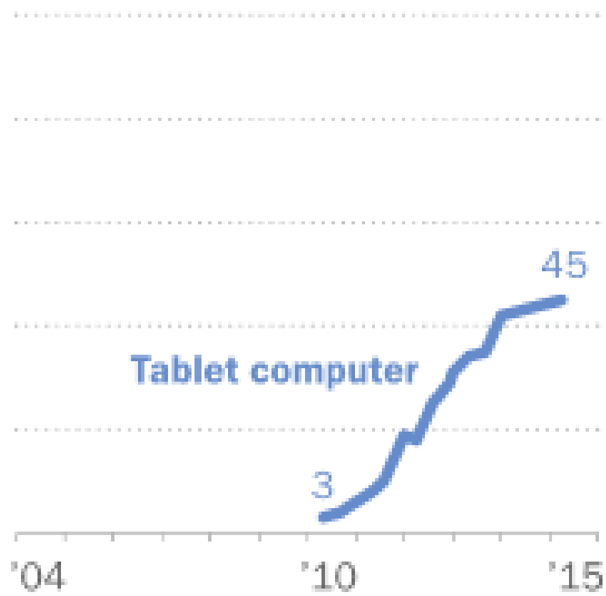

Figure 16: Percentage of American adults who own a tablet

Looking at the Tablet graph provides slightly greater insights than our smartphone graph because Pew Research began collecting data for tablets in 2010. This graph also demonstrates the explosive adoption of digital devices and in the very same year that the decline in checkouts begins.

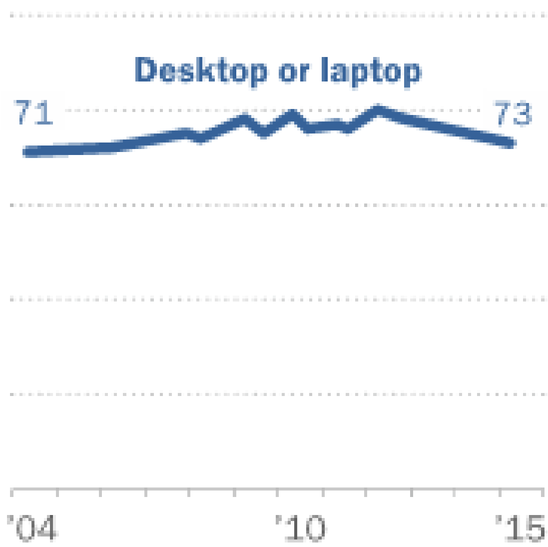

Figure 17: Percentage of American adults who own desktops or laptops

The percentage of ownership for desktops and laptops remained relatively flat from 2004 to 2015. We are not surprised by this as we do not believe laptops to be a driving force in limiting the number of checkouts at the library.

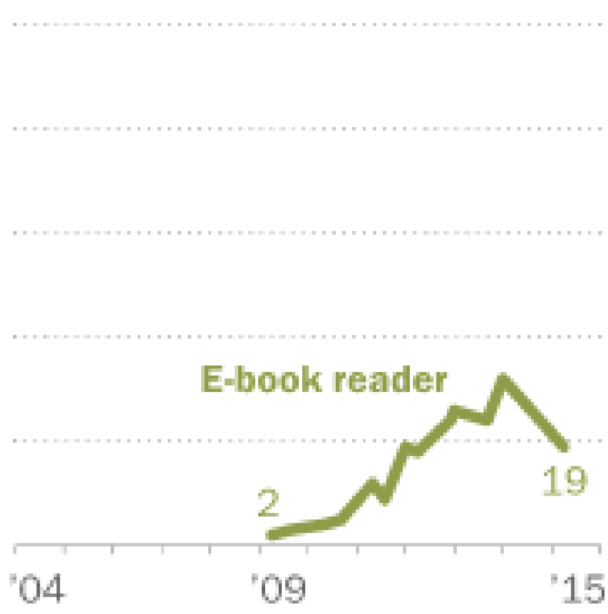

Figure 18: Percentage of American adults who own Ebook readers

We believe this graph to be of extreme importance because we believe it to be one of the main driving forces in the reduction of checkouts since the year 2009. An E-book reader is a version of a tablet that is limited to E-books. An E-book reader allows a user to rent or purchase text from the comfort of their home rather than the local library.

We believe the E-books, smartphones and tablets in combination with the release of digital streaming services to be responisble for the decline in the number of checkouts since 2009. Now, this leads way to numerous applications. A relatively obvious application would apply to publishing companies. Publishing companies should understand the increasing use of digital devices and cater which books they publish physical versions for. For example, textbooks will likely continue to be in demand as they are needed in education. However, they may want to slightly reduce the count of other books and increase the push to digital copies of the book.

We believe another application applies to all writers and content creators. We believe these content creators should attempt to pursue a digital forum rather than a physical one. We believe this pursuit would reach a larger audience and also result in a reduction of publishing costs.

8 FURTHER WORK

Since we did not reach a correlation or any significant relationship between weather and checkout patterns, it might also be a useful idea to gather more dynamic data about the city at any given point. For example, using traffic data in combination with weather data might give us an idea on how weather, impacts traffic, which impacts total library usage.

8.1 Issues and ways we could expand on them in the future

One of the limits with our correlation modeling is that it only looks at the current state of the weather to determine if there is indeed a value between checkout numbers and weather. It would probably be a better idea to expand out of a first degree probability model, to allow past events to help determine if the checkout number will increase or decrease. For example, It is very likely that one hour won't deter someone from going to the library, however, if the morning starts raining, it might deter someone in the evening who planned to visit, to not visit.

Since we were on limited time constraints, we were unable to tackle a few of our initial questions we found interesting relating to these data sets. For example, how the weather may affect different genres or different kinds of checkouts.

8.2 Citations

Anderson, M. (2016, November 17). U.S. Technology Device Ownership 2015. Retrieved from https://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/

Askalidis, Y. (2019, January 8). How and When People use the Public Library. Retrieved from https://towardsdatascience.com/how-and-when-people-use-the-public-library-1b102f58fd8a

Guy, G. B. (2017, May 25). Seattle once again nation's fastest-growing big city; population exceeds 700,000. Retrieved from https://www.seattletimes.com/seattle-news/data/seattle-once-again-nations-fastest-growing-big-city-population-exceeds-700000/