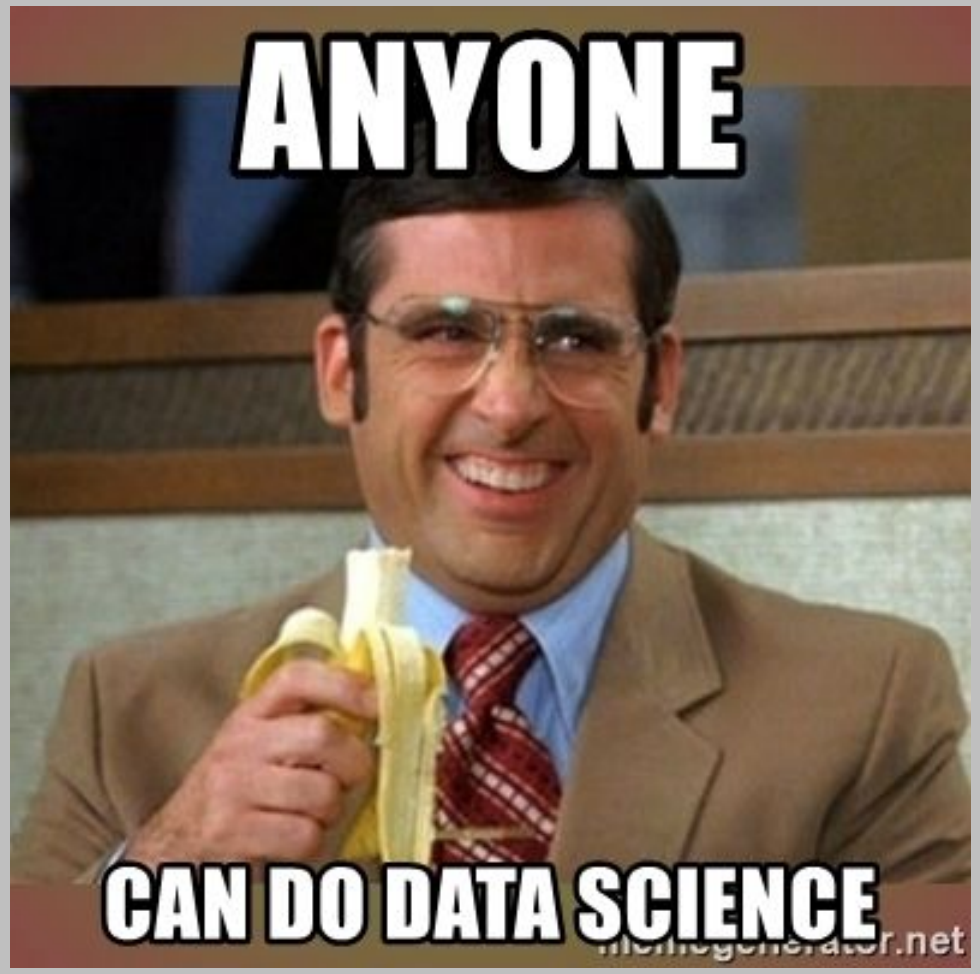
Big Data Analyst: Cohort 1 Graduation

The Struggle Was Real

Welcome To Our Journey



The Meta Data

480
class hours

9
sprints

911
GitHub commits

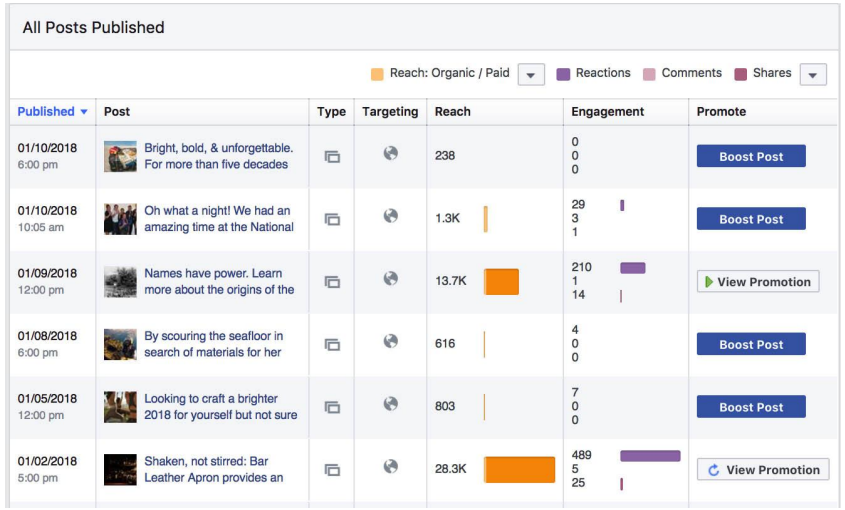
1
machine learning competition

182
pounds of candy consumed (we estimate)

n/a
days the class temp was above 67 degrees

Once upon a time before data class…

Runjini manually copied and pasted data into Excel.



1/10/18				867	20	
10:05 AM				Organic	20	
				867	3	
				Paid	3	
				0	1	
	Oh what a ni	*Photo*	*Public*		1	Boost Post
1/9/18				9.9K	161	
12:00 PM				Organic	161	
				1,019	1	
				Paid	1	
				8,855	9	
	Names have	*Photo*	*Public*		9	View Promot
1/8/18				608	4	
6:00 PM				Organic	4	
				608	0	
				Paid	0	
				0	0	
	By scouring	*Photo*	*Public*		0	Boost Post

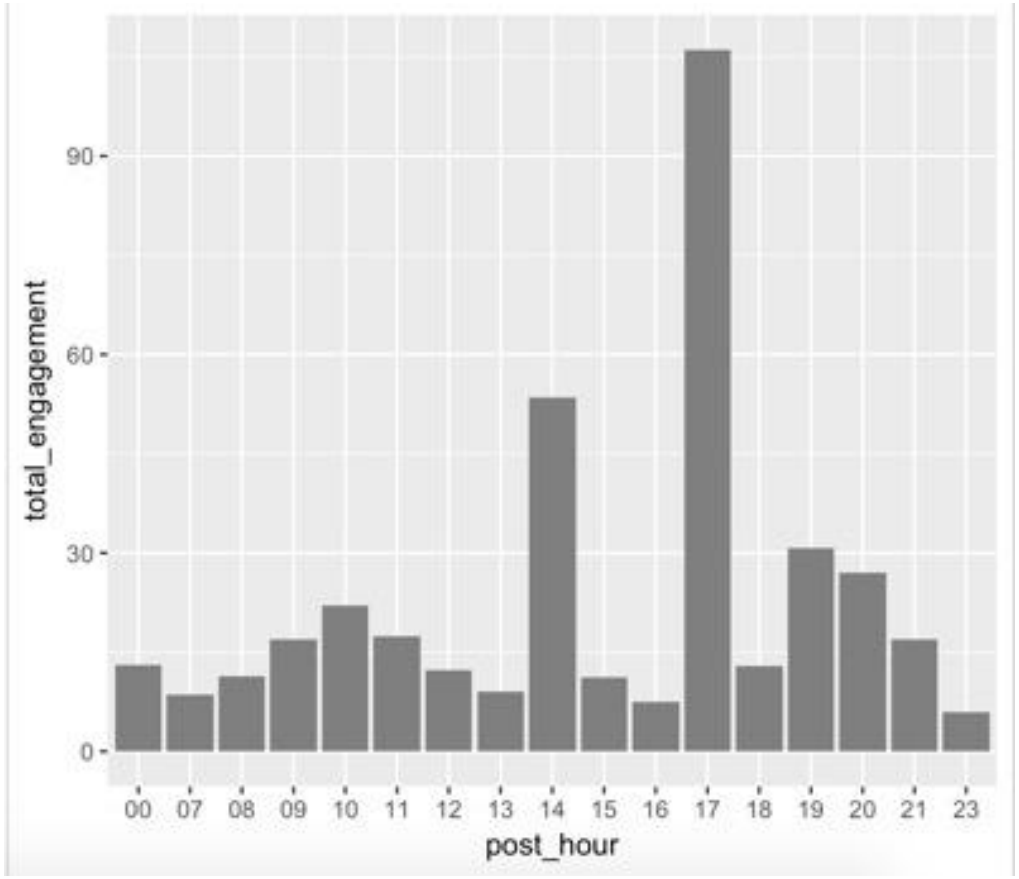
Then we learned about R packages!

19 lines of code replaced tedious copying, pasting, and reformatting.

```
1  install.packages('Rfacebook')
2  library("Rfacebook")
3  myouath <- fbOAuth(app_id='app id goes here',app_secret='app secret goes here')
4  save(myouath, file="myoauth")
5  load("myoauth")
6  me <- getUsers("me", token = myouath)
7  getpagedata <- getPage (page_number_from_website, token = myouath, n = 250)
8  write.csv(getpagedata, "/Users/runjini.murthy/Desktop/GitHub/Runjini_Sprint_4/last-250-wv-fb-posts-Rfacebook.csv")
9  WVFBData <- read.csv(file="/Users/runjini.murthy/Desktop/GitHub/Runjini_Sprint_4/last-250-wv-fb-posts-Rfacebook.csv",
   header=TRUE, sep=",")
10 WVFBData$total_engagement<-WVFBData$likes_count + WVFBData$comments_count + WVFBData$shares_count
11
12 install.packages("parsedate")
13 library(parsedate)
14 WVFBData$parsed_timestamp <- parse_iso_8601(WVFBData$created_time)
15 WVFBData$adjusted_timestamp <- WVFBData$parsed_timestamp - 10*60*60
16 WVFBData$parsed_date <- format(WVFBData$adjusted_timestamp, "%m/%d/%y")
17 WVFBData$parsed_time <- format(WVFBData$adjusted_timestamp, "%H:%M")
18 WVFBData$post_hour <- format(WVFBData$adjusted_timestamp, "%H")
19 write.csv(WVFBData, "/Users/runjini.murthy/Desktop/GitHub/Runjini_Sprint_4/last-250-wv-fb-posts-Rfacebook.csv")
20
21 ggplot(WVFBData) + geom_bar(aes(x=post_hour,y=total_engagement),stat="summary", fun.y = "mean",fill=I("grey50"))
```

And lived happily data after.

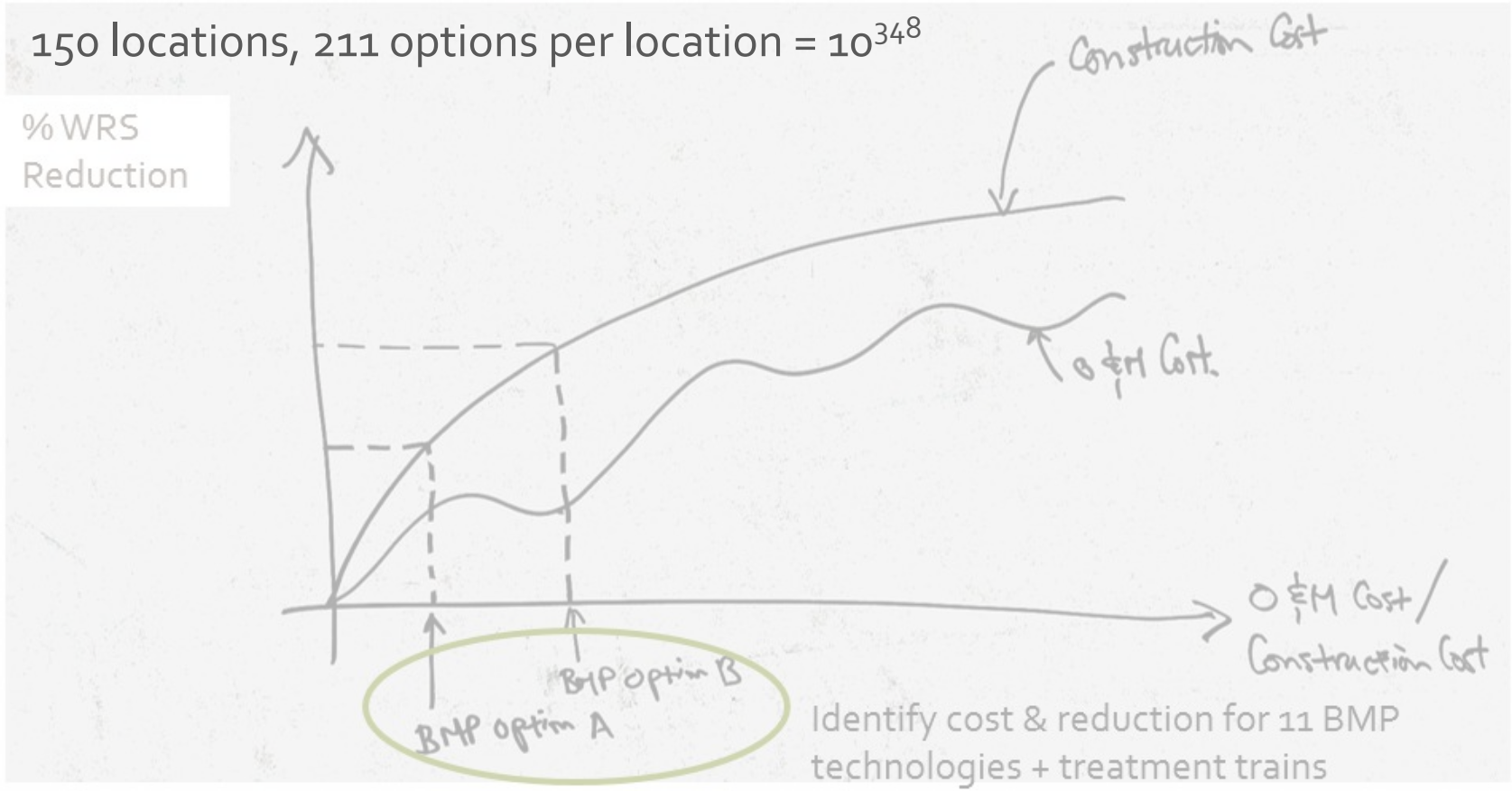
Now it is easier to perform business analysis, like a plot of times to show optimal post hours to drive Facebook engagement.



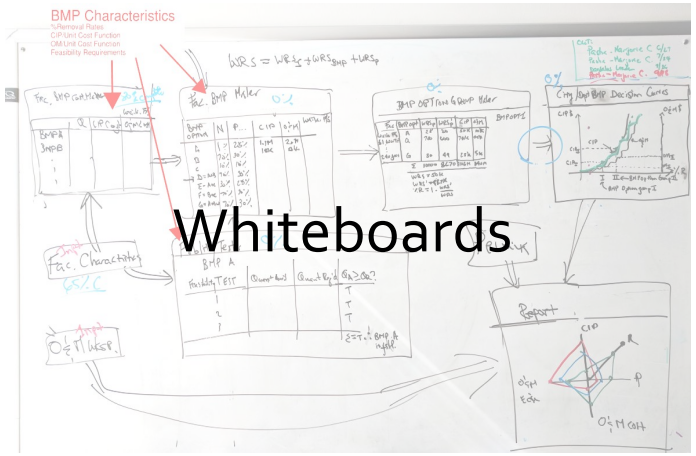
Once upon a time before data class…

Jon asked: How do you analyze 10^{348} options?

150 locations, 211 options per location = 10^{348}



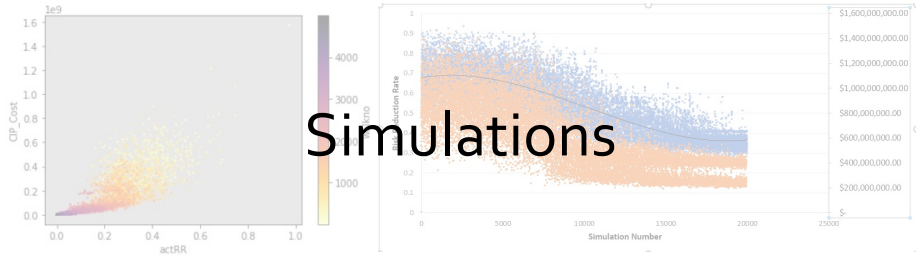
Along the way we met…



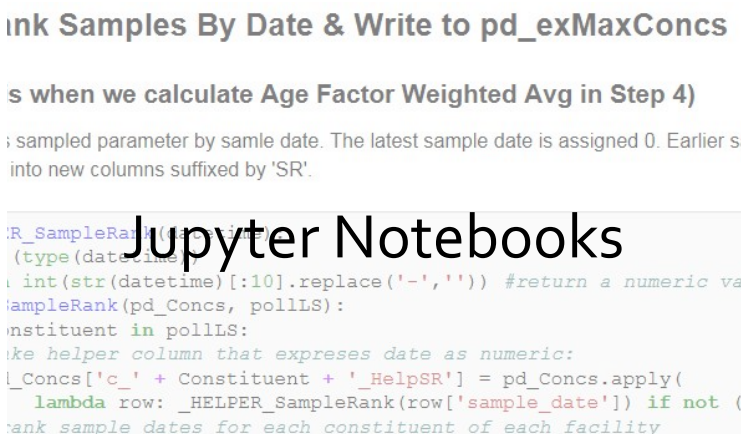
Whiteboards

Facility_ID	id	January	February	March	April	Ma
1	81	3.39	3.85	4.05	2.87	1.8
2	171	4.16	3.02	2.96	1.25	1.6
3	65	4.96	4.13	5.50	2.98	2.5
4	91	5.71	5.46	6.07	4.72	3.4
5	97	5.28	4.98	5.43	3.63	2.8
6	103	3.66	2.01	1.93	1.26	1.0
7	112	3.11	3.53	3.67	2.19	1.3
8	121	5.25	4.90	5.96	4.11	3.3

Pandas



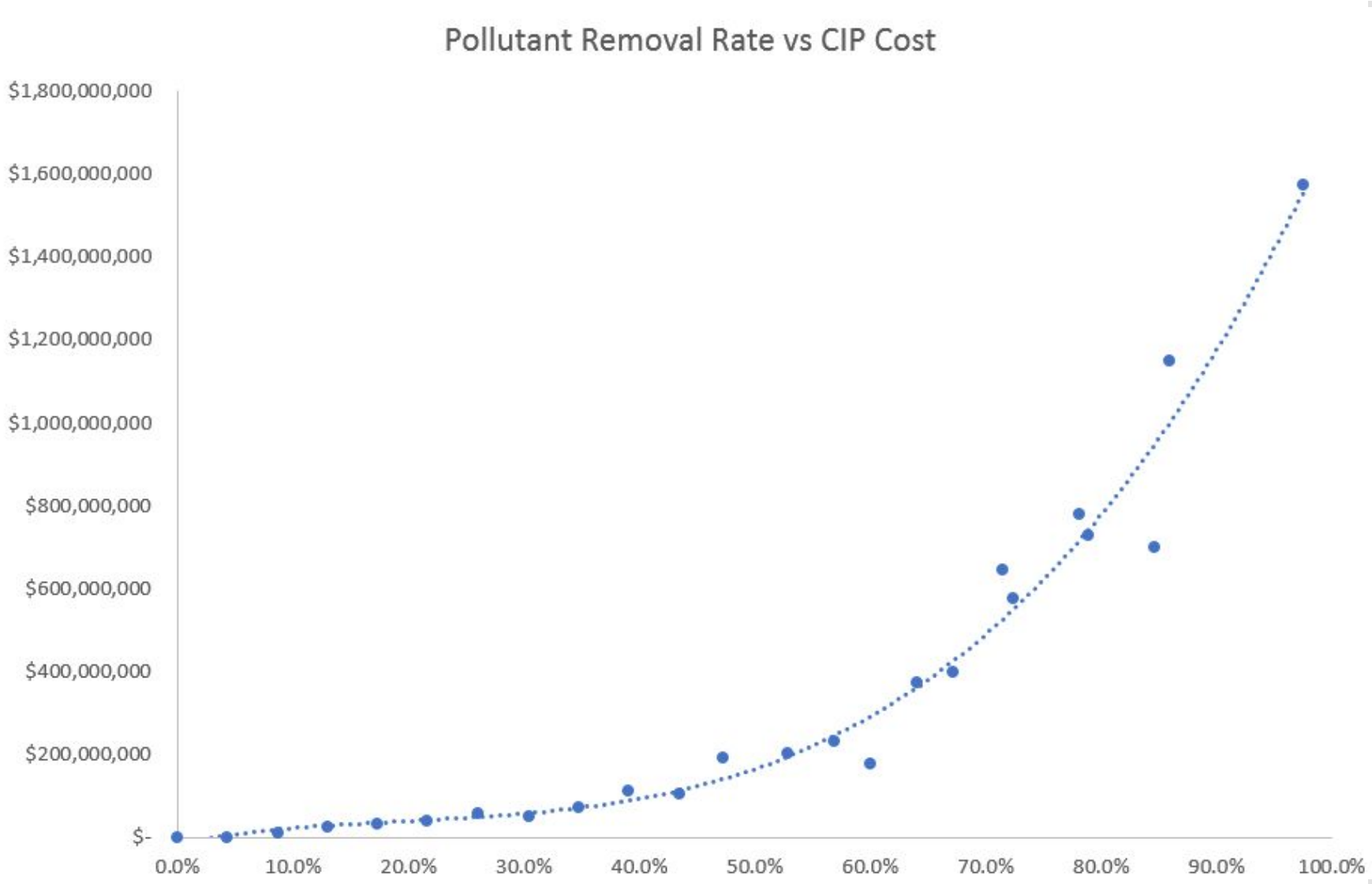
Simulations



Jupyter Notebooks

Cloud Computing

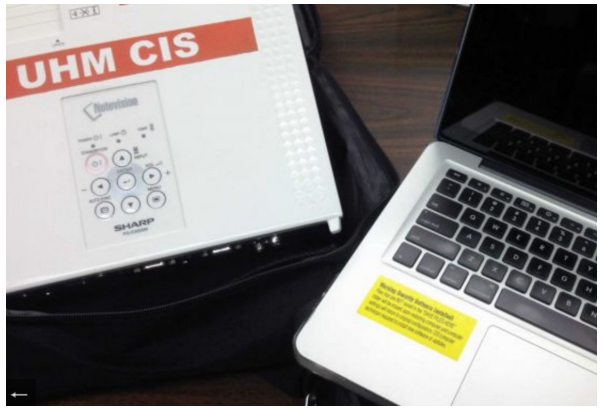
And lived happily data after.



Once upon a time before data class…

For Tori, it all started in the 9th grade with Ms. Carlston…

Fast forward to Tori's sophomore year of college where she takes her first HTML class.



Fast forward to January 2016. Girl meets boy. Girl meets Devleague.

The love-hate relationship continues to grow…

And lived happily data after.

```
Victorias-MacBook:~ victorialarson$ ls
Applications            Library                 Public
Desktop                 Movies                  US_diseases.R
Documents               Music                   USdisease_questions.R
Downloads               Pictures                anaconda
```

```r
#getting page 7 from the file
get_file_from_date <- function(start_date) {
  filename <- paste(start_date, ".pdf", sep = "")
  file_text <- pdf_text(filename)
  # Want to focus on page 3 of the vector in the list.
  page7 <- file_text[[7]]
  #Using regexpr to find the location of RevPAR\n
  tindex <- regexpr("Running\nTop 25 Markets", page7)
  #deleting "Running\nTop 25 Markets" and everything before it using substr
  page7 <- substring(page7,tindex+23)
  # String split with "\n" - creates new lines
  newline <- strsplit(page7, "\n")
  #using substring turned my page into a list, I need it to be a character
  newchar <- unlist(newline)
  #taking out all of the rows I dont need.
  newchar <- newchar[-(1)]
  newchar <- newchar[-(26:28)]
  # Extracted all of the floats from the vectors
  extract <- (str_extract_all(newchar,"[+-]?([0-9]*[.])?[0-9]+"))
  # turning the list back into a character
  un_list <- as.numeric(unlist(extract))
  # telling it there need to be 18 lines
  rep_list <- rep(1:18, times=length(un_list)/18)
  # I dont know whats going on here
  split_list <- split(un_list, rep_list)
  # making the columns
  df <- cbind.data.frame(split_list, stringsAsFactors=F)
  # taking out the columns that I dont want
  df <- df[-(8:18)]
  #adding column names
  row.names(df) <- c("Anaheim/SantaAna_CA","Atlanta_GA","Boston_MA","Chicag
```

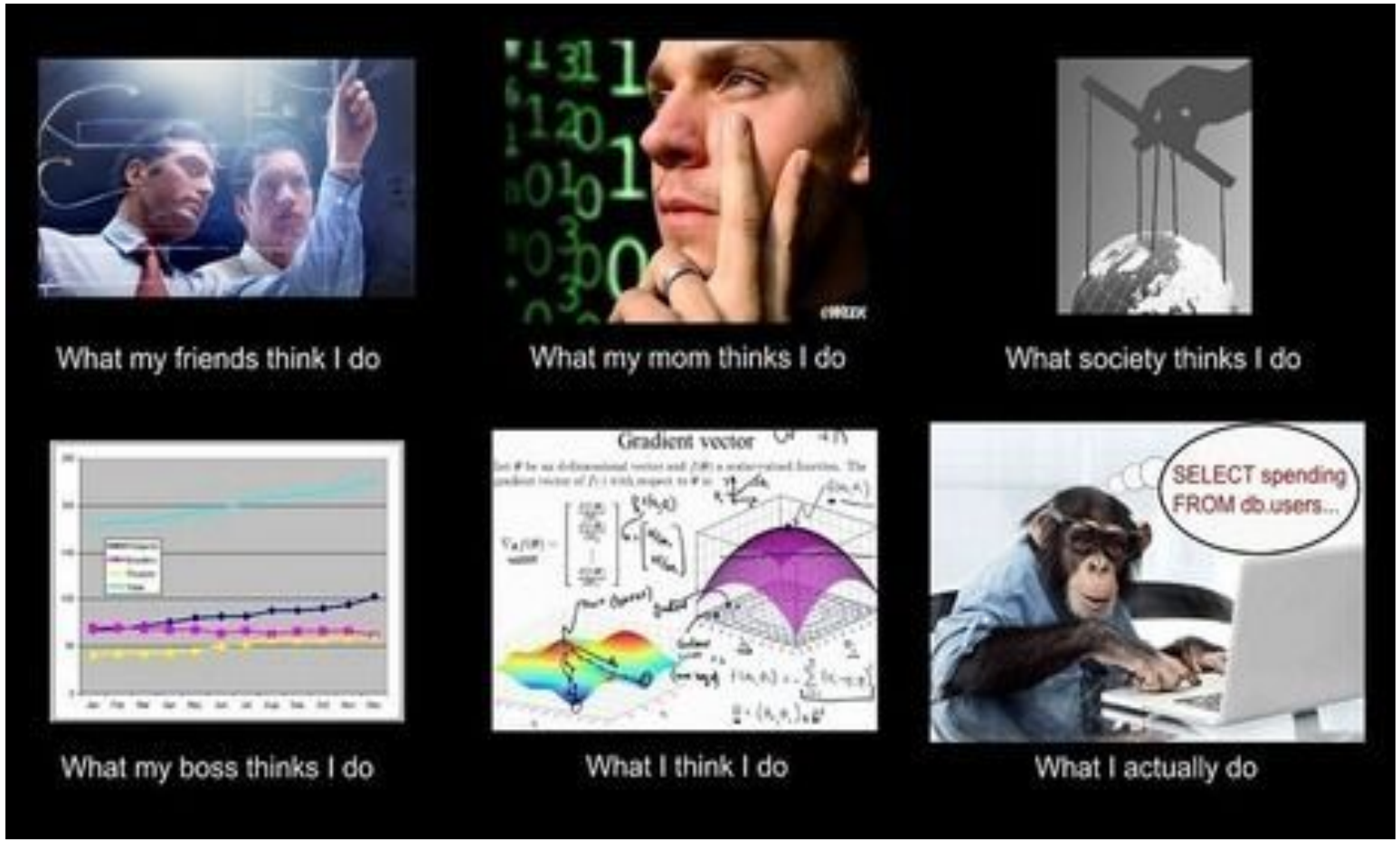
A Data Offering for DevLeague

DATA PREP	Intro To Data and Mathematical Foundations
MODEL 1	Data Analytics, Data Collection & The Analytical Process
MODEL 2	Business Statistics & Probability Theory with Microsoft Excel
MODEL 3	Exploratory Data Analysis & Business Intelligence Reporting with Tableau
MODEL 4	Introduction to SQL Databases & Relational Data Modeling
MODEL 5	Introduction to Programming in R & R Studio

Oh wait, there's more!

MODEL 6	Python Programming & Introduction to Big Data Analysis
MODEL 7	Introduction to Machine Learning & Predictive Modeling
MODEL 8	Advanced Data Visualization & Storytelling
MODEL 9	Data Governance & Security in The Enterprise
CAPSTONE PROJECT	The Big Reveal

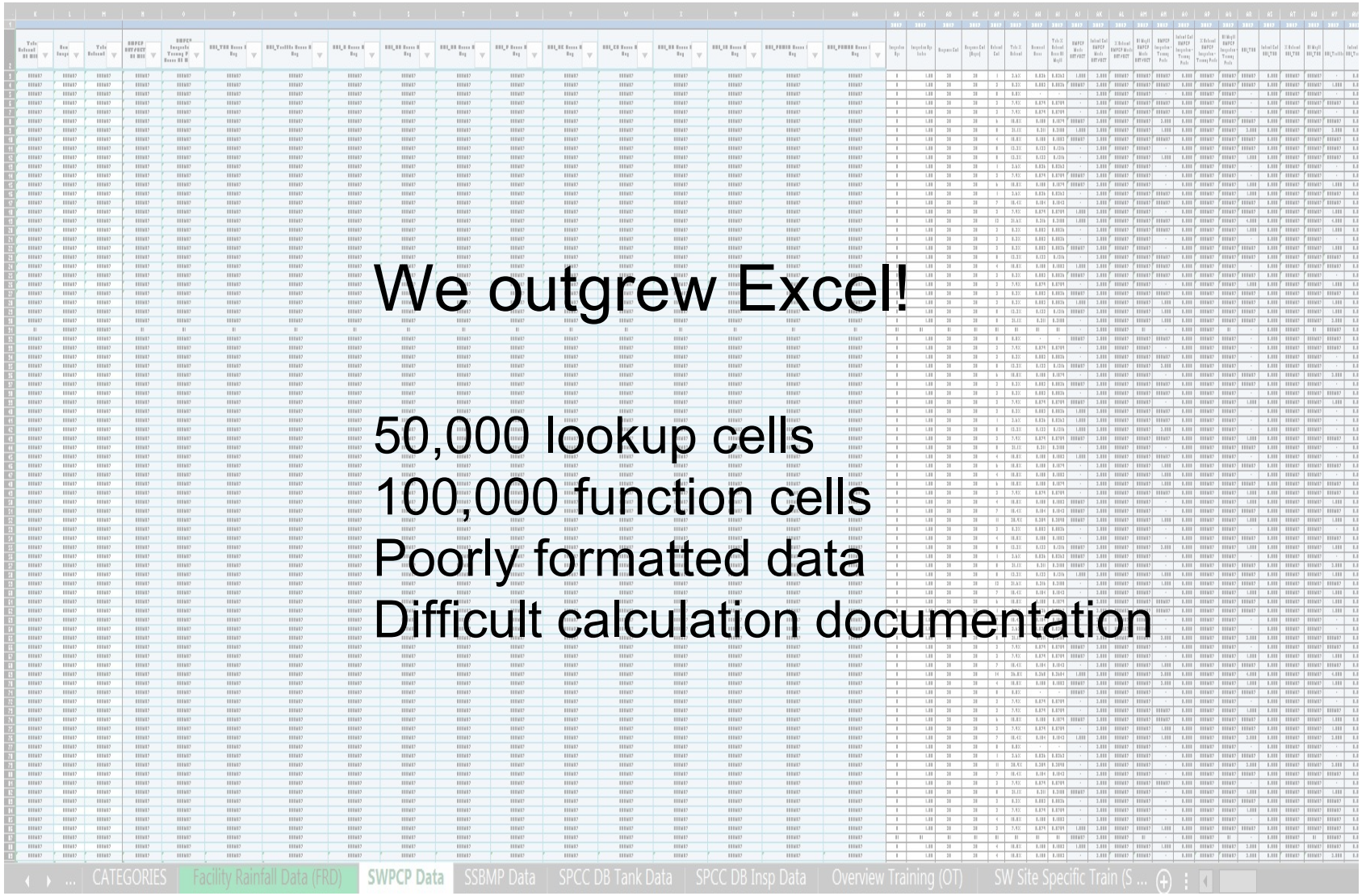
Where were we for the last 40 weeks?



Haiku:
An Ode to Data Analysis

Google is your friend.
Wrangling takes most of the time.
One typo ruins all.

Why can't we just use Excel?

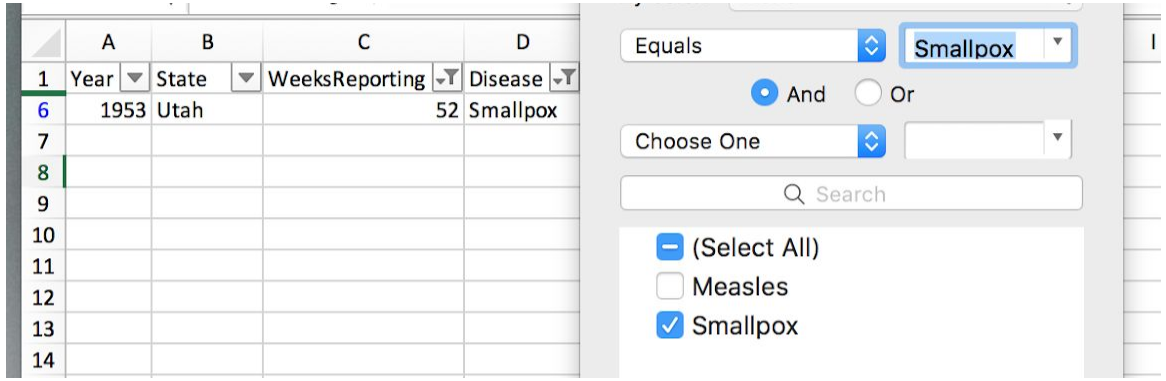


We outgrew Excel!

50,000 lookup cells
100,000 function cells
Poorly formatted data
Difficult calculation documentation

What's the deal with R vs. Python?

Excel



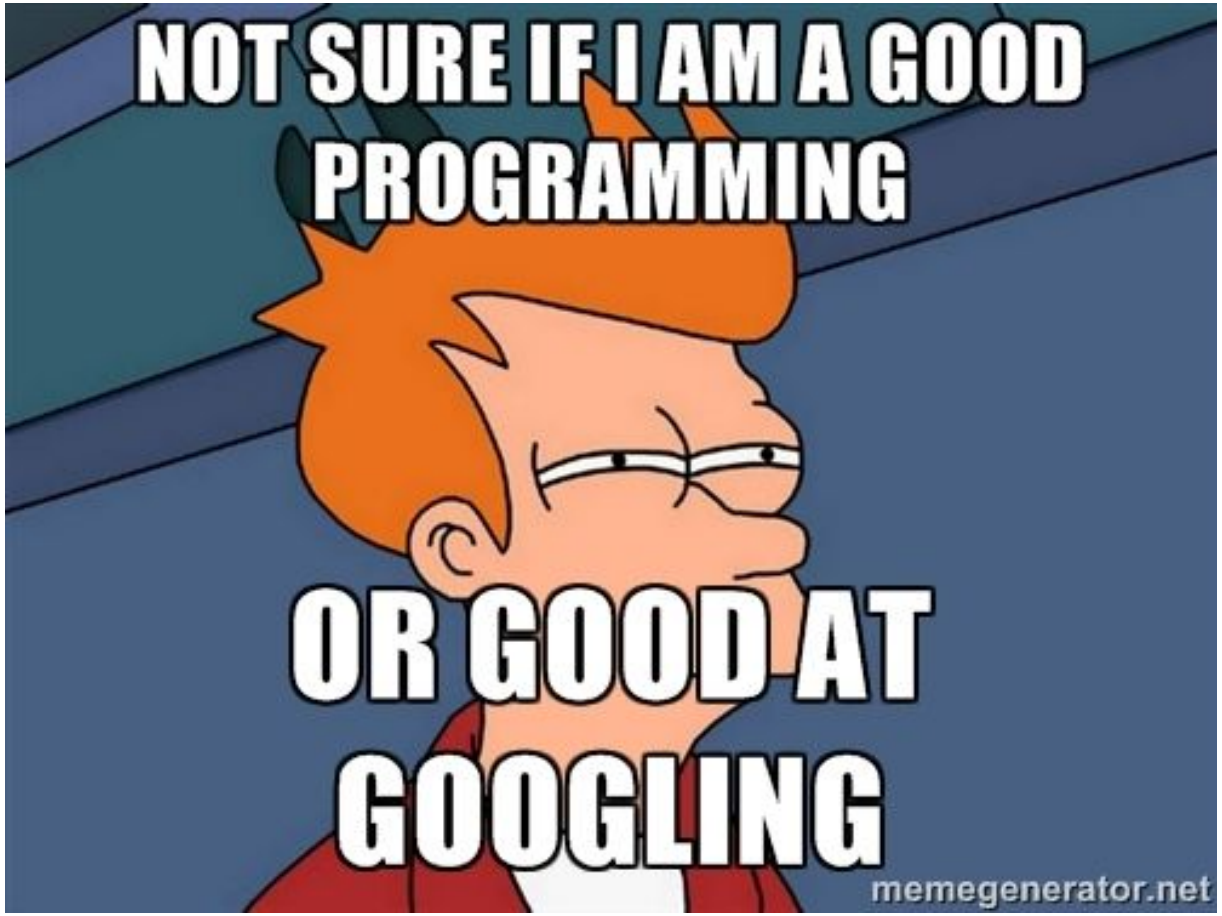
R

```r
#filtering out just Small pox  from the weeks reporting 52
Smlpx52wks <- subset(Week52_reporting, disease=="Smallpox")
```

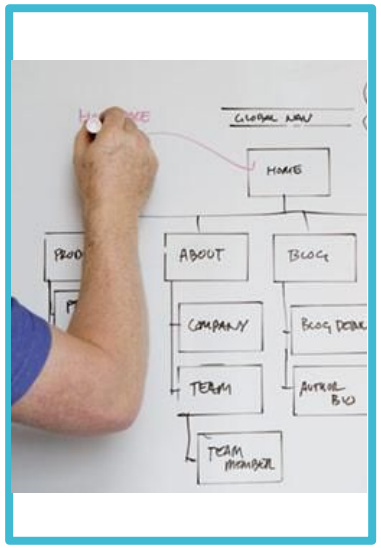
Python

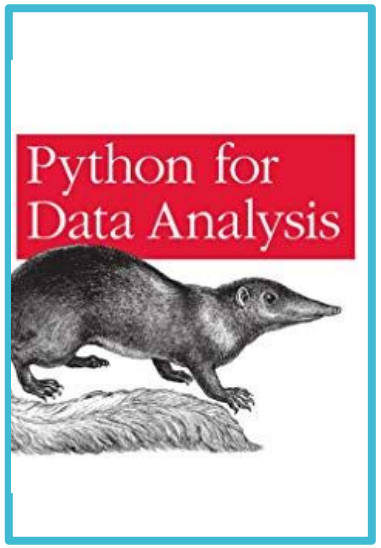
```python
Smlpx52wks = Week52_reporting.loc[Week52_reporting['disease'] == 'Smallpox']
```

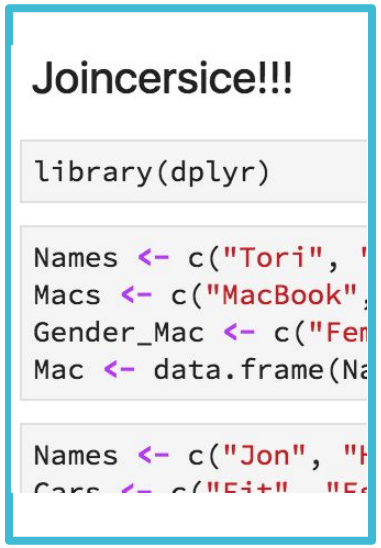
How We Learned



Helpful Resources



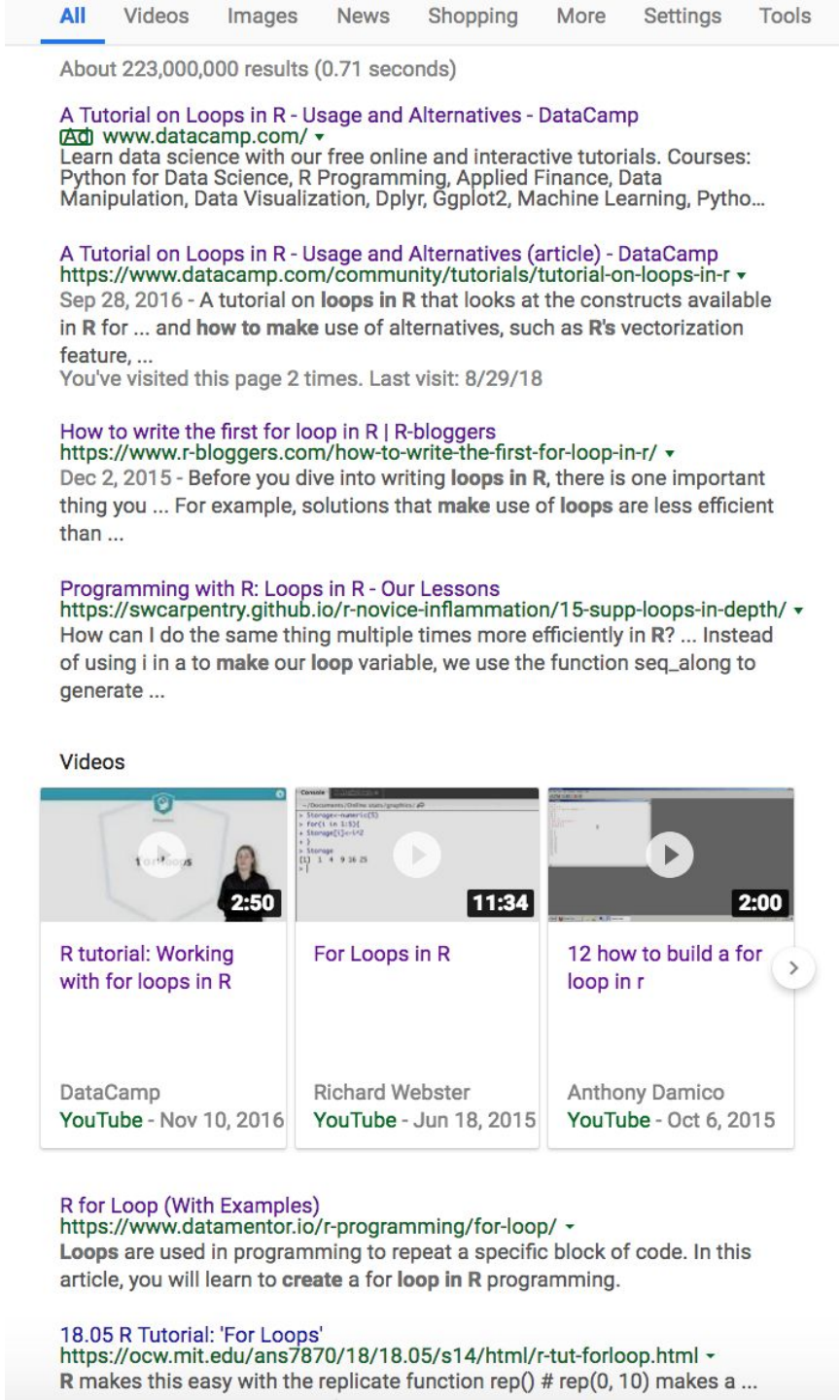




Group Projects

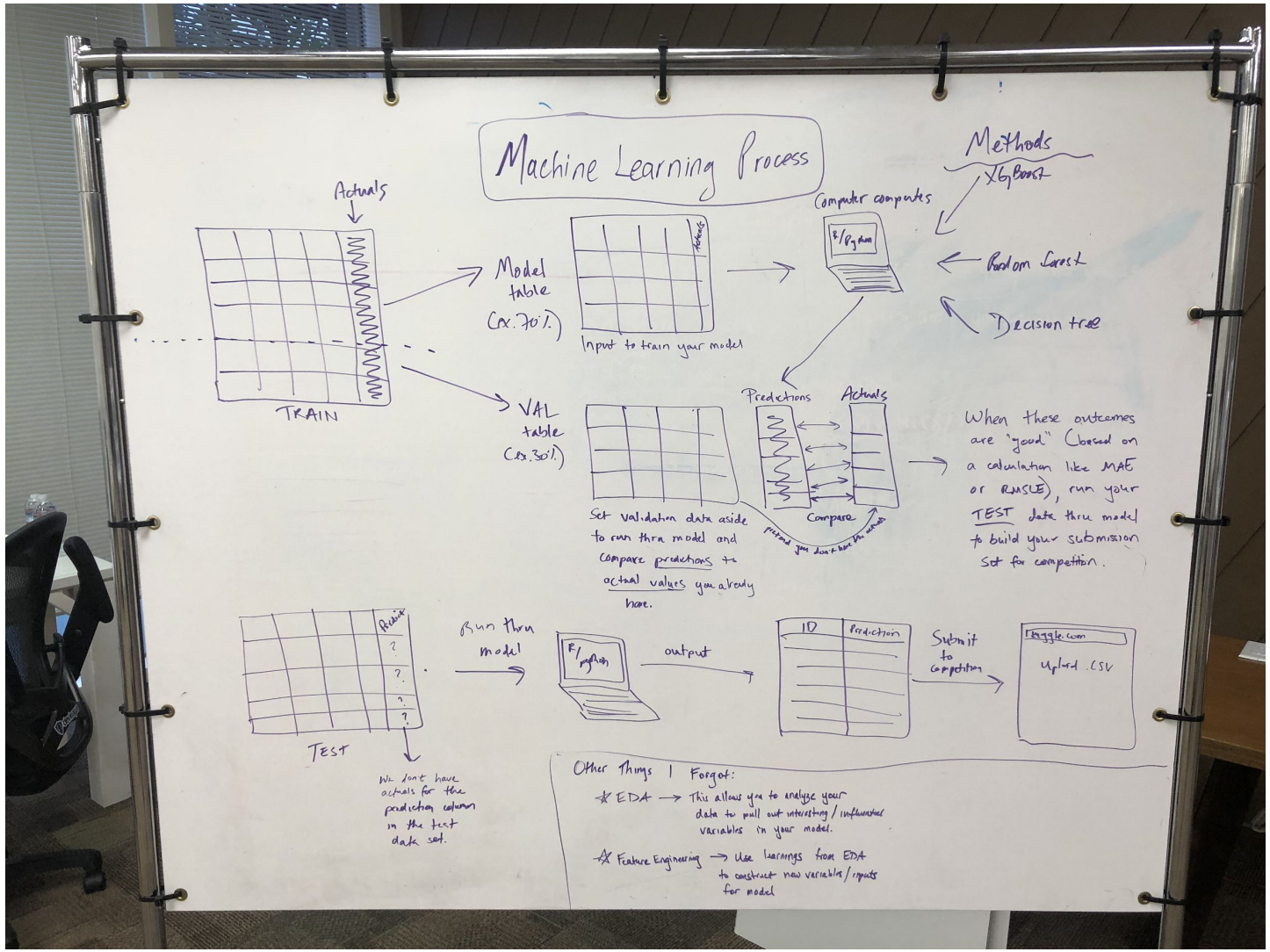
Helpful Resources

Googling Like Ninjas



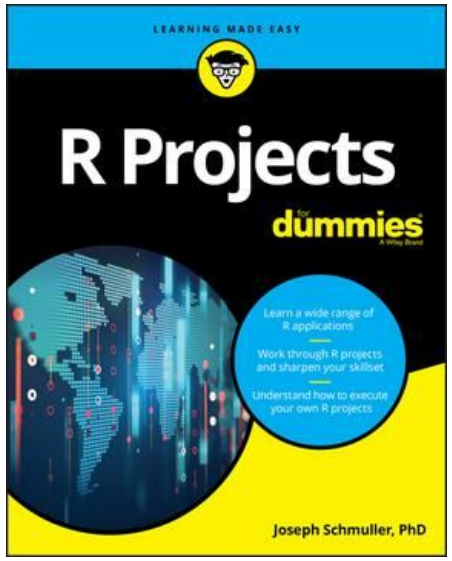
Helpful Resources

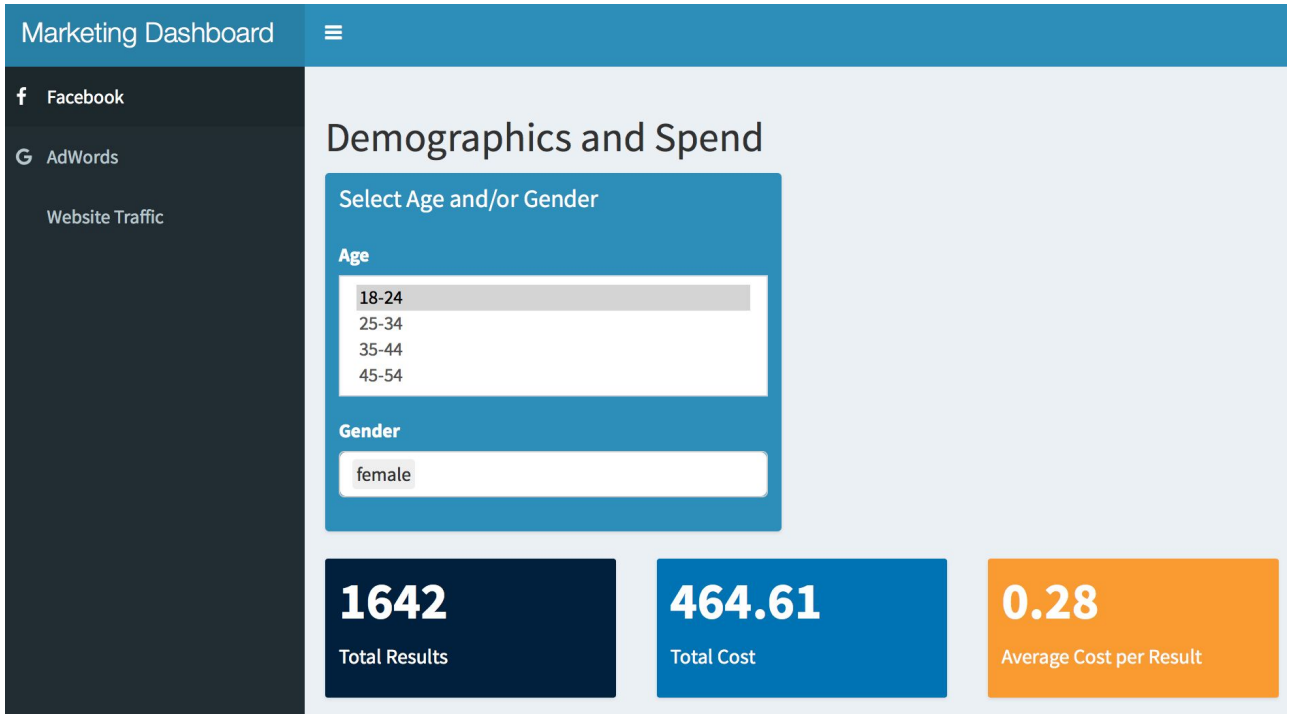
Whiteboarding Like Champs



Helpful Resources

Reading Like It's Our Job





EDA. WTF.

DEFINE the acronym

Exploratory **D**ata **A**nalysis

What it **IS**

Learning more about your data without making assumptions

An iterative process

What it **IS NOT**

Complete visualizations with fancy aesthetics

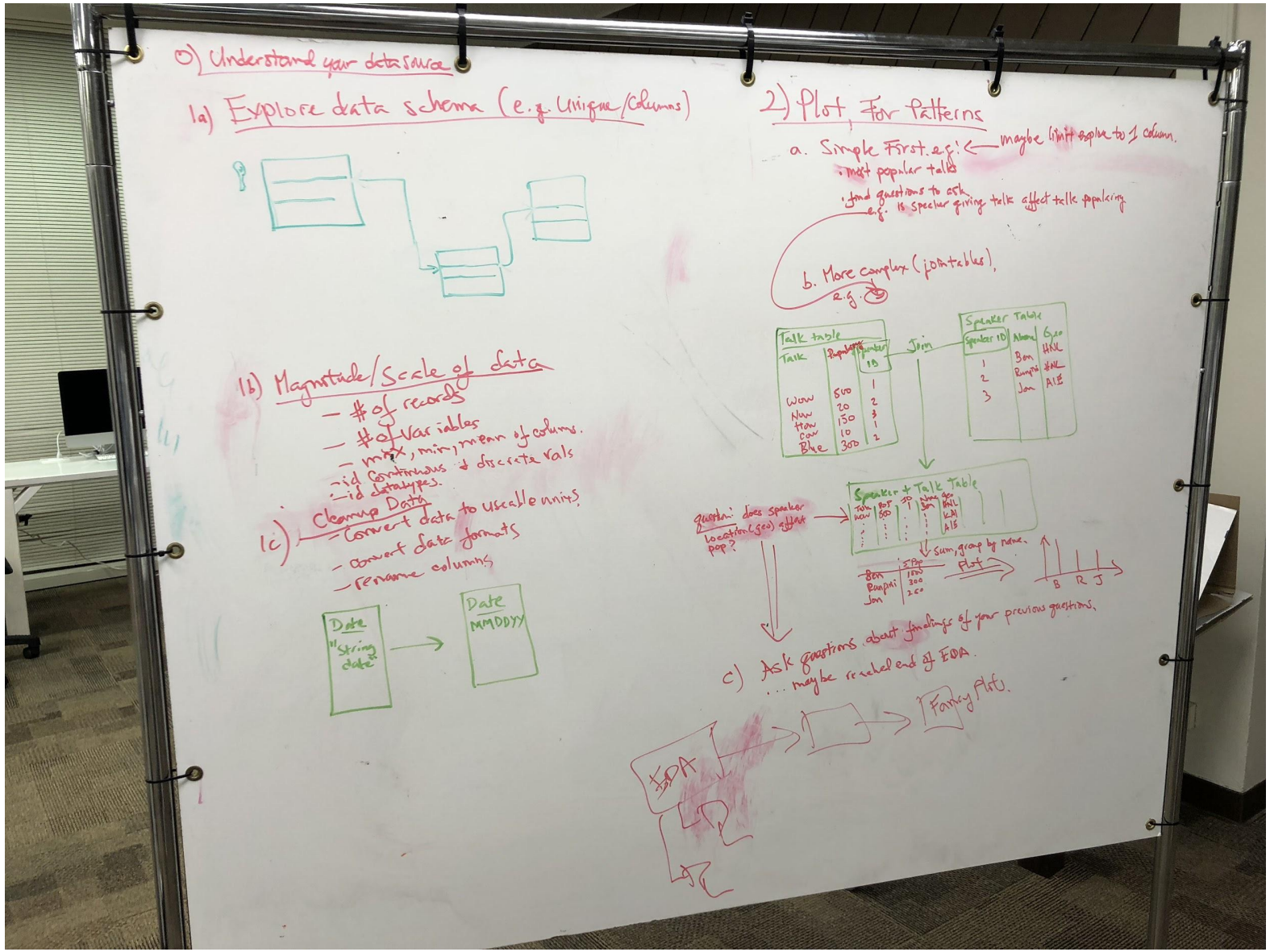
HOW it is done

Summary statistics

Quick visualizations (bar plots, histograms, scatterplots)

Example EDA Process

1. Understand data source
2. Explore data schema
3. What is the magnitude/scale of data?
4. Clean up data
5. Plot for patterns: simple → complex
6. Ask more questions
7. Do it again and again



What are ceRcizes?

We created "ceRcizes" as a teaching and learning tool.

Joincersice!!!

In [ ]:
```
library(dplyr)
```

In [43]:
```
Names <- c("Tori", "Ben", "Hunter", "Runjini")
Macs <- c("MacBook", "Pro2014", "Pro2017", "Pro2011" )
Gender_Mac <- c("Female", "Male", "Male", "Female")
Mac <- data.frame(Names, Macs, Gender_Mac)
```

In [44]:
```
Names <- c("Jon", "Hunter", "Runjini")
Cars <- c("Fit", "Escape", "Mini")
Gender_Car <- c("Male", "Male", "Female")
Car <- data.frame(Names, Cars, Gender_Car)
```

In [49]:
```
Mac
Car
```

Names	Macs	Gender_Mac
Tori	MacBook	Female
Ben	Pro2014	Male
Hunter	Pro2017	Male
Runjini	Pro2011	Female

Exciting Things We Did and Learned

Runjini: What I Learned

SQL

SELECT *

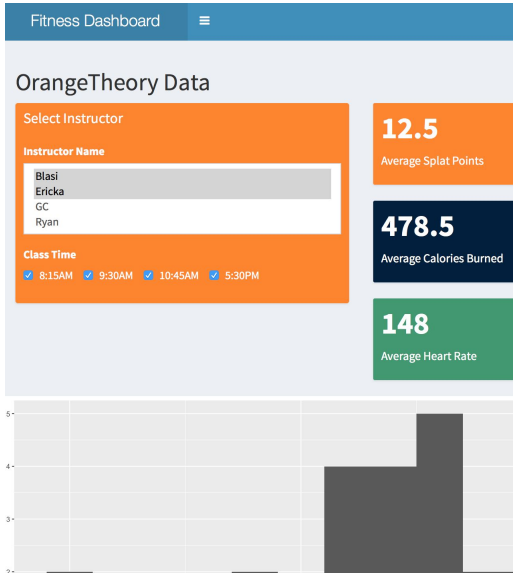
FROM data_class

WHERE googled = TRUE

RFacebook

+

Shiny



SQL + Wine

Did you know… 21% of wines on the Kaggle data set of over 130,000 wines are described using the phrases “oak” or “wood”?

```
SELECT COUNT(description)
FROM wine_reviews
WHERE description like '%oak%' OR description like '%wood%'
```

Also researched:

- How many countries are represented on this list? **44**
- How many wines have a perfect score of 100? **19**
- What is the average points rating? **88.44**

SQL + Wine

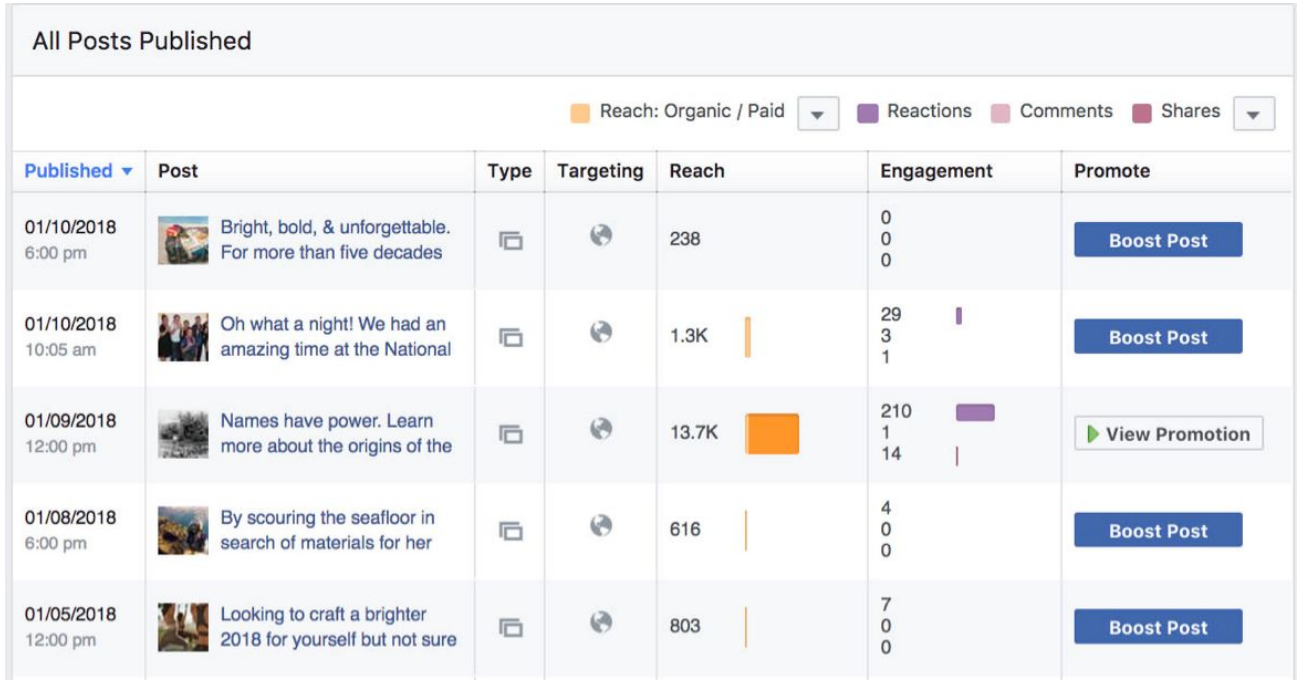
Provide all information on wines from India.

```
SELECT *
FROM wine_reviews
WHERE country = 'India'
```

Total rows loaded: 9

	wine_review_ID	country	description	designation	points	price	province
1	6673	India	The nose carries notes of cassis, blackberry and bell pepper. An approachable, yet co...	Dindori Reserve	93	20	Nashik
2	29302	India	A nose of cut herbs and just-mown grass backed by notes of white peach barely prep...		89	12	Nashik
3	37225	India	Pineapple, grapefruit, and apricot show brightly on the nose and palate. Simultaneousl...		90	10	Nashik
4	56903	India	Aromas of blackberry, cherry preserves, white chocolate and a hint of green pepper p...	Estate Bottled	91	12	Nashik
5	83622	India	Pineapple, grapefruit, and apricot show brightly on the nose and palate. Simultaneousl...		90	10	Nashik
6	83634	India	Aromas of pink grapefruit, grass and coriander seed open the palate of this wine. Pass...		90	12	Nashik
7	88134	India	Dark violet-red in color, this wine has a bouquet of blueberry, brambles and flint. It is s...	Estate Bottled	90	13	Nashik
8	108251	India	Deep garnet in the glass, this wine has a nose of blackberry and roasting meat. It offer...	Dindori Reserve	92	19	Nashik
9	126932	India	This wine features a fresh nose of grapefruit and lemon peel. It is bright on the palate, ...		87	12	Nashik

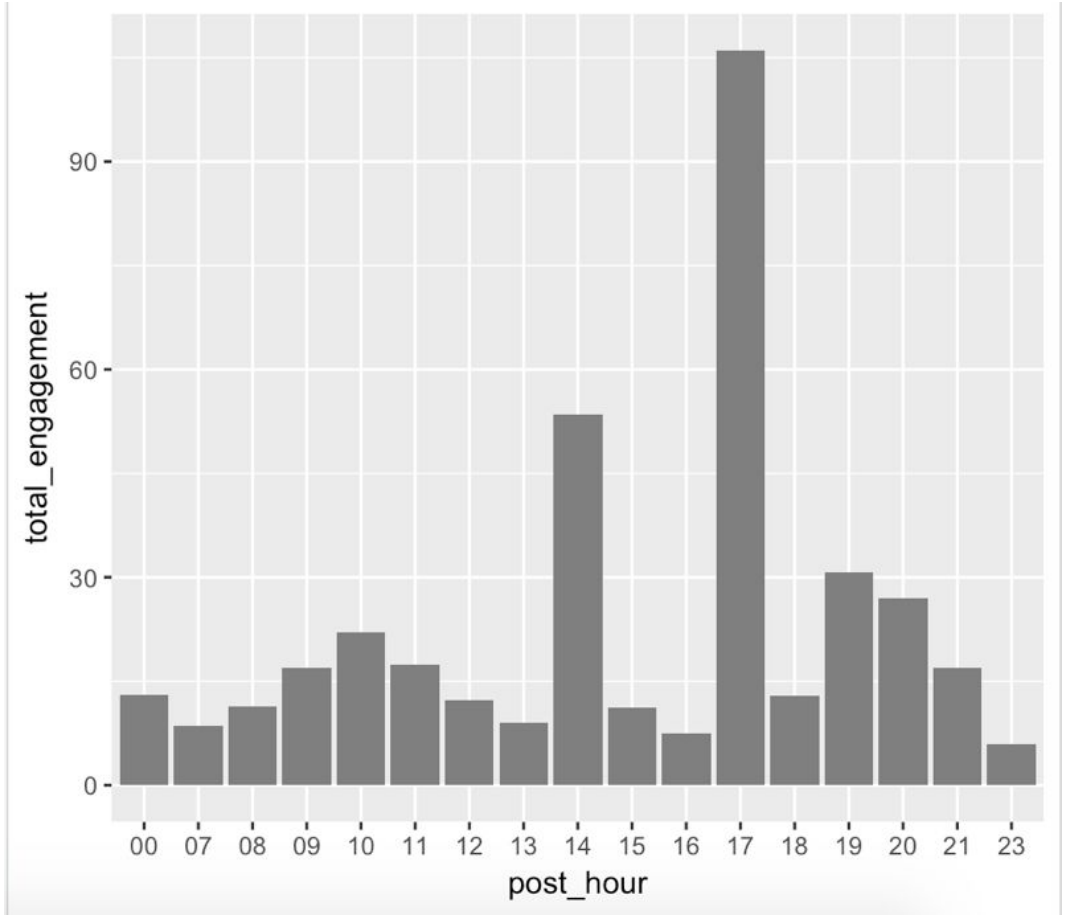
R + Facebook



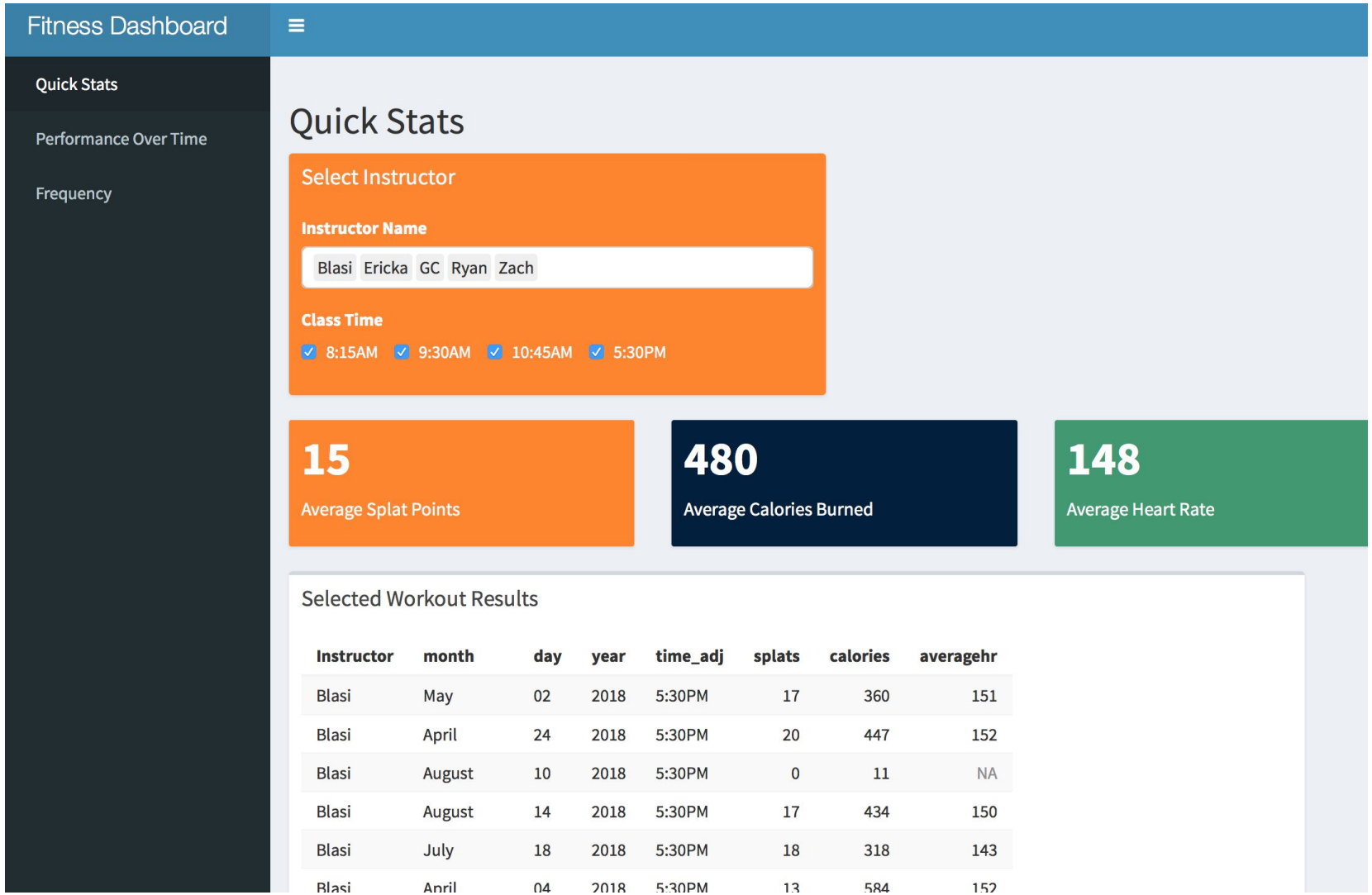
from_id	from_name	message	created_time	type	link	id
3.106683e+14	Ward Village	Where Hawaii name...	2016-10-26T22:45:...	link	http://bit.ly/2ePPUyV	310668279073583_...
3.106683e+14	Ward Village	Have you heard? The...	2016-10-27T22:45:...	link	http://bit.ly/2eA87ST	310668279073583_...
3.106683e+14	Ward Village	We know how busy l...	2016-10-28T22:15:...	link	http://bit.ly/2fohLJn	310668279073583_...
3.106683e+14	Ward Village	Happy Halloween fro...	2016-10-31T22:55:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	It's official! We've su...	2016-11-01T22:00:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	Will you be joining u...	2016-11-02T22:45:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	Coming to Ward Vill...	2016-11-05T05:03:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	Last year, the Ice Rin...	2016-11-08T04:10:...	link	http://bit.ly/2fudoKb	310668279073583_...
3.106683e+14	Ward Village	Next week, we'll be ...	2016-11-09T02:54:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	Grand Opening of S...	2016-11-11T19:34:...	video	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	One of our favorite t...	2016-11-16T07:52:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	Mahalo to everyone ...	2016-11-16T23:31:...	video	https://vimeo.com/...	310668279073583_...
3.106683e+14	Ward Village	Today marks such a ...	2016-11-17T18:51:...	video	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	*NA*	2016-11-19T01:55:...	photo	https://www.facebo...	310668279073583_...
3.106683e+14	Ward Village	This week marked 2 ...	2016-11-19T18:00:...	photo	https://www.facebo...	310668279073583_...

R + Facebook

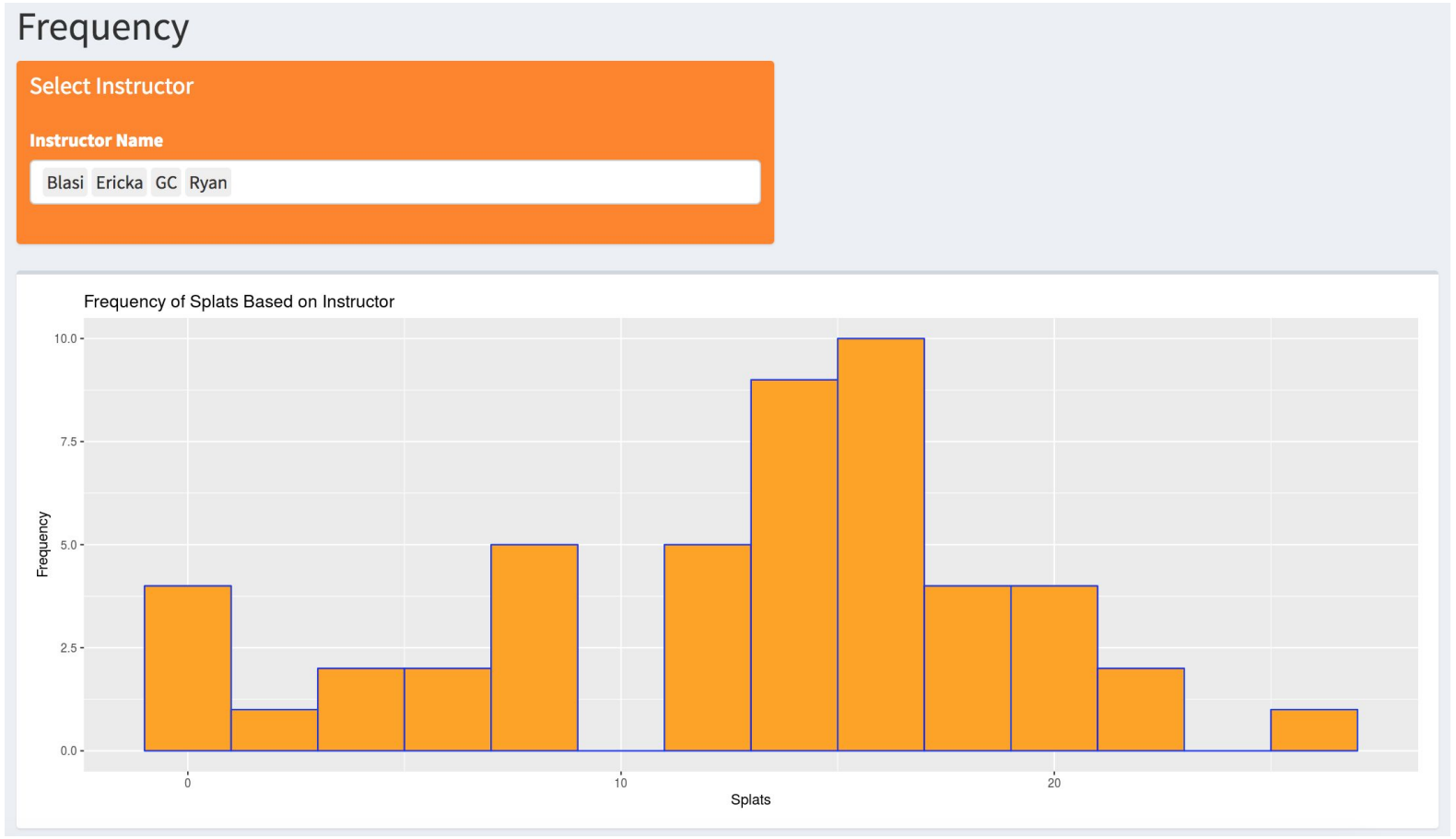
Using RFacebook, I could determine the best times for us to post in terms of which hours would yield the highest total engagement: 5pm, 2pm, and 7pm.



Shiny


Fitness Dashboard
Quick Stats
Performance Over Time
Frequency
Quick Stats
Select Instructor
Instructor Name
Blasi Ericka GC Ryan Zach
Class Time
8:15AM 9:30AM 10:45AM 5:30PM
15
Average Splat Points
480
Average Calories Burned
148
Average Heart Rate
Selected Workout Results
Instructor month day year time_adj splats calories averagehr
Blasi May 02 2018 5:30PM 17 360 151
Blasi April 24 2018 5:30PM 20 447 152
Blasi August 10 2018 5:30PM 0 11 NA
Blasi August 14 2018 5:30PM 17 434 150
Blasi July 18 2018 5:30PM 18 318 143
Blasi April 04 2018 5:30PM 13 584 152

Shiny


Frequency
Select Instructor
Instructor Name
Blasi
Ericka
GC
Ryan
Frequency of Splats Based on Instructor
10.0
7.5
5.0
2.5
0.0
Frequency
0
10
20
Splats

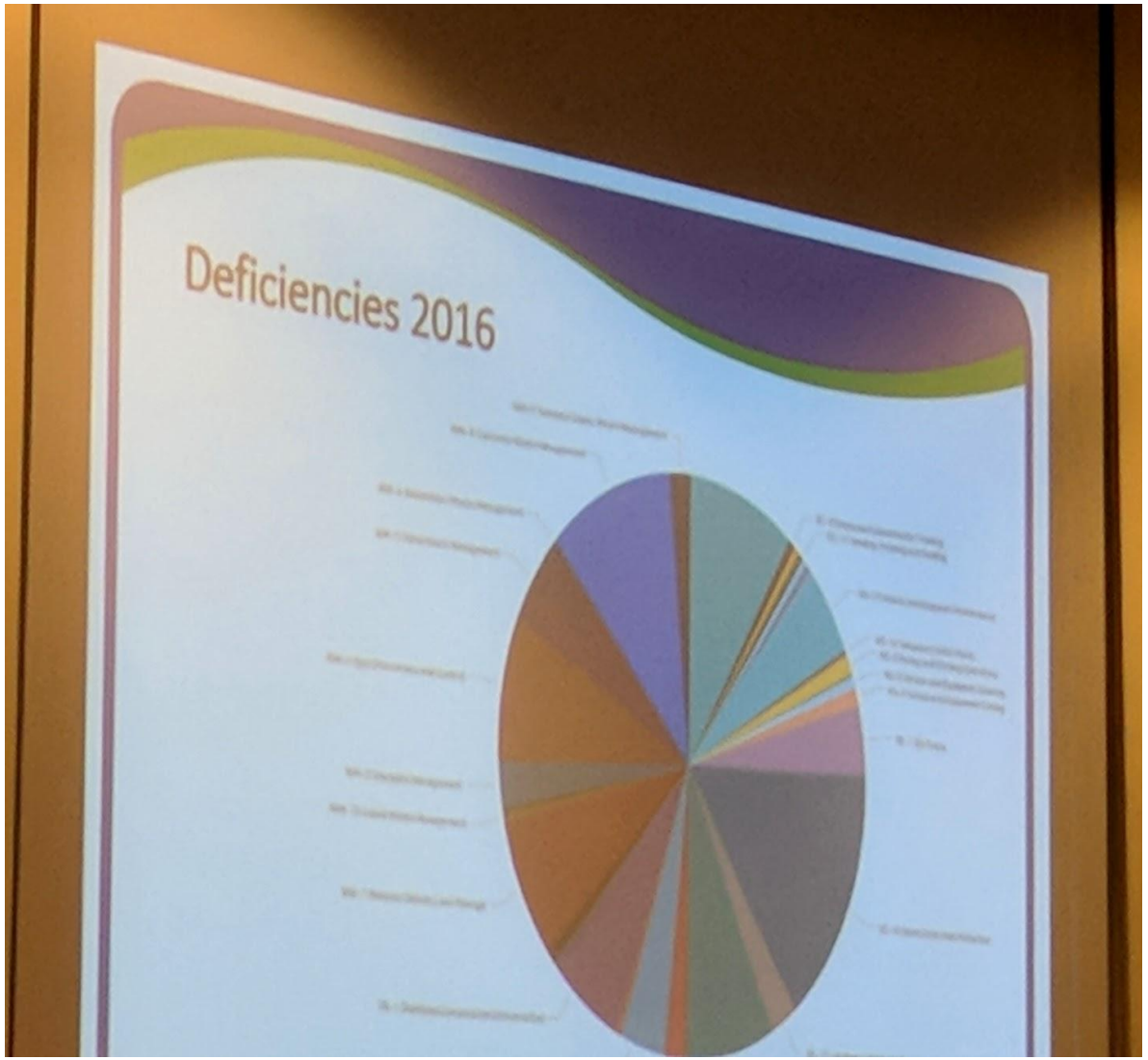
Jon: What I Learned

Too Much Pie!

Stories on Jupyter

Auto-Tagging Email

Too much pie!



Stories on Jupyter

	I	P	Q	R	S	T
1						
2	Facility Name (Appendix A1)	NEL_TSS Score AF Wtd Avg	NEL_Turbidity Score AF Wtd Avg	NEL_N Score AF Wtd Avg	NEL_NN Score AF Wtd Avg	NEL_AN Score AF Wtd Avg
3	Ahuimanu Dewatering Facility	0.011	0.011	-	-	-
4	Aiea Fire Station	0.132	0.132	0.011	0.011	0.011
5	Aikahi Fire Station	-	-	-	-	-
6	Aircraft One Fire Station	-	-	-	-	-
7	Airport Vehicle Maintenance Yard	0.031	0.031	-	-	-
8	Ala Moana Regional Park (including Maintenance Baseyard)	0.007	0.007	-	-	-
9	Ala Wai Golf Course	0.472	0.472	0.493	0.493	0.493
10	Alapai Police Station	0.124	0.124	-	-	-
11	Central Fire Station	0.015	0.015	-	-	-
12	Central Oahu Regional Park (including Maintenance Baseya	0.085	0.085	-	-	-
13	East Kapolei Fire Station	0.004	0.004	-	-	-
14	Ewa Beach Fire Station	0.004	0.004	-	-	-
15	Ewa Refuse Convenience Center	0.160	0.160	-	-	-

Stories on Jupyter

Jupyter Notebooks offer a multi-purpose solution to tell a story.

Description

Step 4.c: Calculate Age Factor Weighted (AFW) Average Exceedances

We claim that more recent samples better represent current housekeeping and samplir
necessary to represent long term trends and to dampen whipsaw effects that may occu
weighted average exceedance(AFW average exceedance) considers these two viewpc
on older samples. The equation is:

$$AFW_p = \frac{\sum C_p e^{-SRn}}{\sum e^{-SRn}}$$

Coding

```
12      df[AFColName + '_AF'] = np.exp(-df[AFColName + '_DR']) #calculate each year's weig
13      df[AFColName + '_AF*Val'] = df[AFColName + '_AF']*df[AFColName] #apply weight to i
14      #for each facility, add af weights and wt applied values up. enter to temp datafra
15      tmp = df.groupby(['Facility_Name'])[AFColName + '_AF', AFColName + '_AF*Val'].sum(
16          columns = {AFColName + '_AF':AFColName + '_AFSUM',AFColName + '_AF*Val':AFColN
17  #     display(tmp)
18      df=pd.merge(df,tmp,on='Facility_Name',how='inner') #merge
```

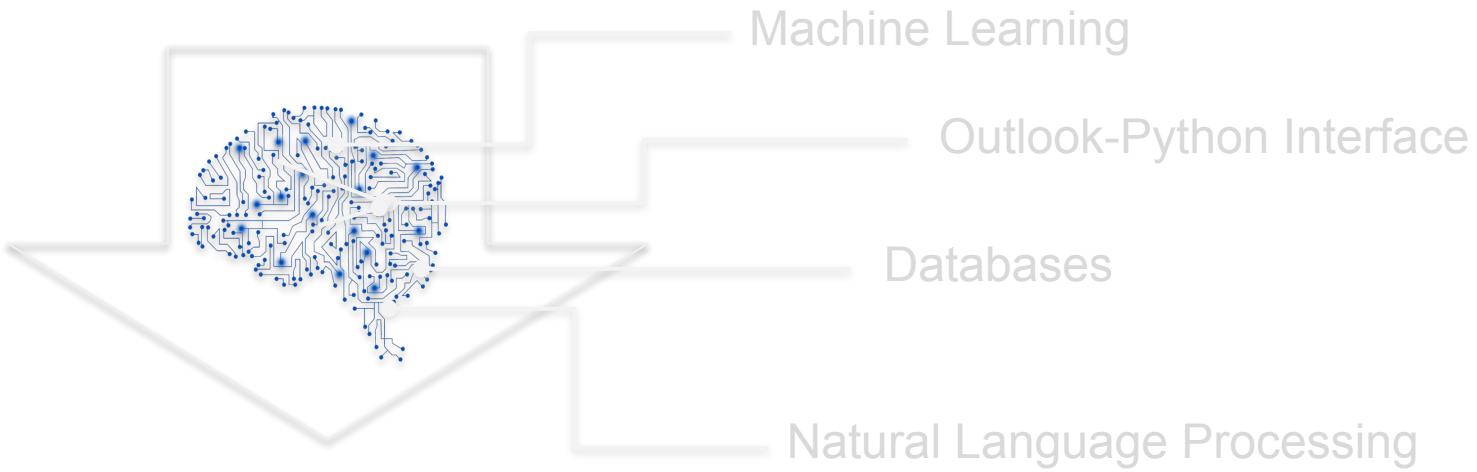
Results

Fiscal Year	Deficient_Count	Total_Count	Deficiency_Rate	Deficiency_Rate_AF
2013	1	38	0.026316	0.018316
2014	0	38	0.000000	0.049787
2015	3	38	0.078947	0.135335
2016	1	38	0.026316	0.367879
2017	1	38	0.026316	1.000000

Categorize Email

Use previously categorized email…

FROM	SUBJECT	CATEGORIES
Underground Solutions	Comparing HDD to Open-Cut	Meeting
EnviroCert International, Inc.	Hurricane Florence: EnviroCert East Coast Office Closure	Flyers
Storm Water Solutions	Earn PDH credits without the travel expenses | SWS Webinar Fest	Flyers
Cortana	Heads-up	Meeting

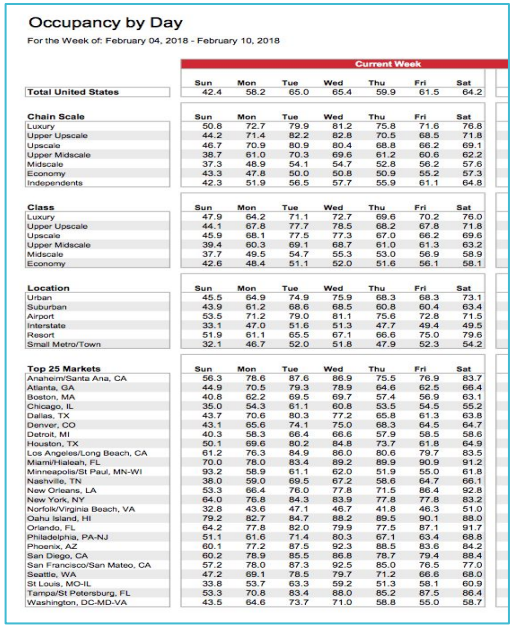


To categorize new email.

FROM	SUBJECT	CATEGORIES
Lauren Baltas, Storm Water…	Hurricane Florence threatens the East Coast	
Jeremy Huckaby	Statement of Clarification	
Weeklysafety.com, LLC	NEW OFFER! Fast 100 Now Includes Construction AND General I…	

Tori: What I Learned

PDF Tools

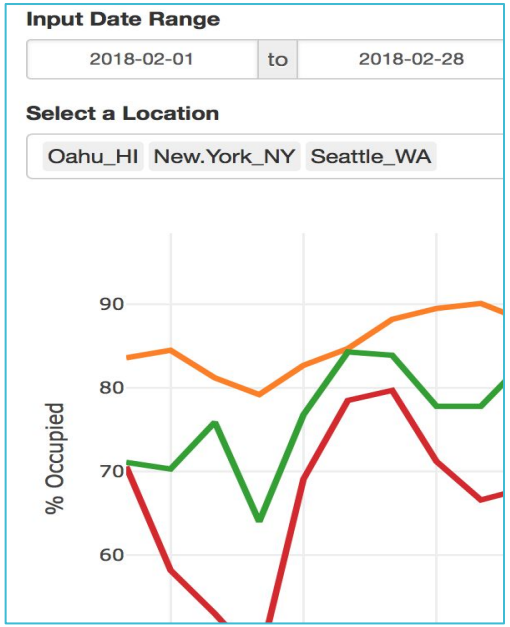


Occupancy by Day

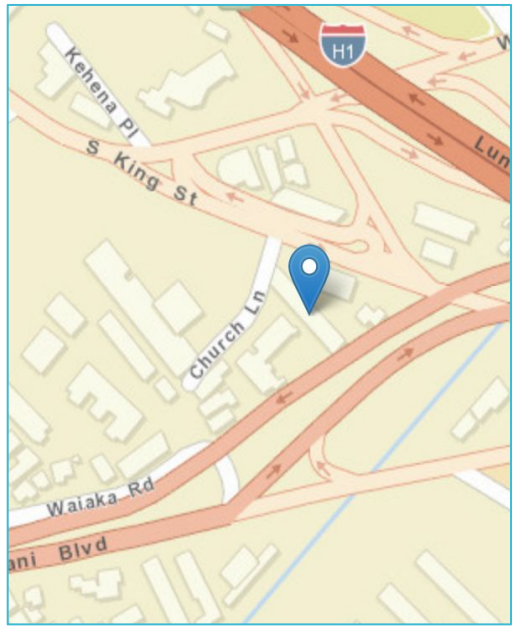
For the Week of: February 04, 2018 - February 10, 2018

	Current Week						
	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Total United States	42.4	58.2	65.0	65.4	59.9	61.5	64.2
Chain Scale							
Luxury	50.8	72.7	79.9	81.2	75.8	71.6	76.8
Upper Upscale	44.2	71.4	82.2	82.8	70.5	68.5	71.8
Upscale	46.7	70.9	80.9	80.4	68.8	66.2	69.1
Upper Midscale	38.7	61.0	70.3	69.6	61.2	60.6	62.2
Midscale	37.3	48.9	54.1	54.7	52.8	56.2	57.6
Economy	43.3	47.8	50.0	50.8	50.9	55.2	57.3
Independents	42.3	51.9	56.5	57.7	55.9	61.1	64.8
Class							
Luxury	47.9	64.2	71.1	72.7	69.6	70.2	76.0
Upper Upscale	44.1	67.8	77.7	78.5	68.2	67.8	71.8
Upscale	45.9	68.1	77.5	77.3	67.0	66.2	69.6
Upper Midscale	39.4	60.3	69.1	68.7	61.0	61.3	63.2
Midscale	37.7	49.5	54.7	55.3	53.0	56.9	58.9
Economy	42.6	48.4	51.1	52.0	51.6	56.1	58.1
Location							
Urban	45.5	64.9	74.9	75.9	68.3	68.3	73.1
Suburban	43.9	61.2	68.6	68.5	60.8	60.4	63.4
Airport	53.5	71.2	79.0	81.1	75.6	72.8	71.5
Interstate	33.1	47.0	51.6	51.3	47.7	49.4	49.5
Resort	51.9	61.1	65.5	67.1	66.6	75.0	79.6
Small Metro/Town	32.1	46.7	52.0	51.8	47.9	52.3	54.2
Top 25 Markets							
Anaheim/Santa Ana, CA	56.3	78.6	87.6	86.9	75.5	76.9	83.7
Atlanta, GA	44.9	70.5	79.3	78.9	64.6	62.5	68.4
Boston, MA	40.8	62.2	69.5	69.7	57.4	56.9	63.1
Chicago, IL	35.0	54.3	61.1	60.8	53.5	54.5	55.2
Dallas, TX	43.7	70.6	80.3	77.2	65.8	61.3	63.8
Denver, CO	43.1	65.6	74.1	75.0	68.3	64.5	64.7
Detroit, MI	40.3	58.3	66.4	66.6	57.9	58.5	58.6
Houston, TX	50.1	69.6	80.2	84.8	73.7	61.8	64.9
Los Angeles/Long Beach, CA	61.2	76.3	84.9	86.0	80.6	79.7	83.5
Miami/Hialeah, FL	70.0	78.0	83.4	89.2	89.9	90.9	91.2
Minneapolis/St Paul, MN-WI	93.2	58.9	61.1	62.0	51.9	55.0	61.8
Nashville, TN	38.0	59.0	69.5	67.2	58.6	64.7	66.1
New Orleans, LA	53.3	66.4	76.0	77.8	71.5	86.4	92.8
New York, NY	64.0	76.8	84.3	83.9	77.8	77.8	83.2
Norfolk/Virginia Beach, VA	32.8	43.6	47.1	46.7	41.8	46.3	51.0
Oahu Island, HI	79.2	82.7	84.7	88.2	89.5	90.1	88.0
Orlando, FL	64.2	77.8	82.0	79.9	77.5	87.1	91.7
Philadelphia, PA-NJ	51.1	61.6	71.4	80.3	67.1	63.4	68.8
Phoenix, AZ	60.1	77.2	87.5	92.3	88.5	83.6	84.2
San Diego, CA	60.2	78.9	85.5	86.8	78.7	79.4	88.4
San Francisco/San Mateo, CA	57.2	78.0	87.3	92.5	85.0	76.5	77.0
Seattle, WA	47.2	69.1	78.5	79.7	71.2	66.6	68.0
St Louis, MO-IL	33.8	53.7	63.3	59.2	51.3	58.1	60.9
Tampa/St Petersburg, FL	53.3	70.8	83.4	88.0	85.2	87.5	86.4
Washington, DC-MD-VA	43.5	64.6	73.7	71.0	58.8	55.0	58.7

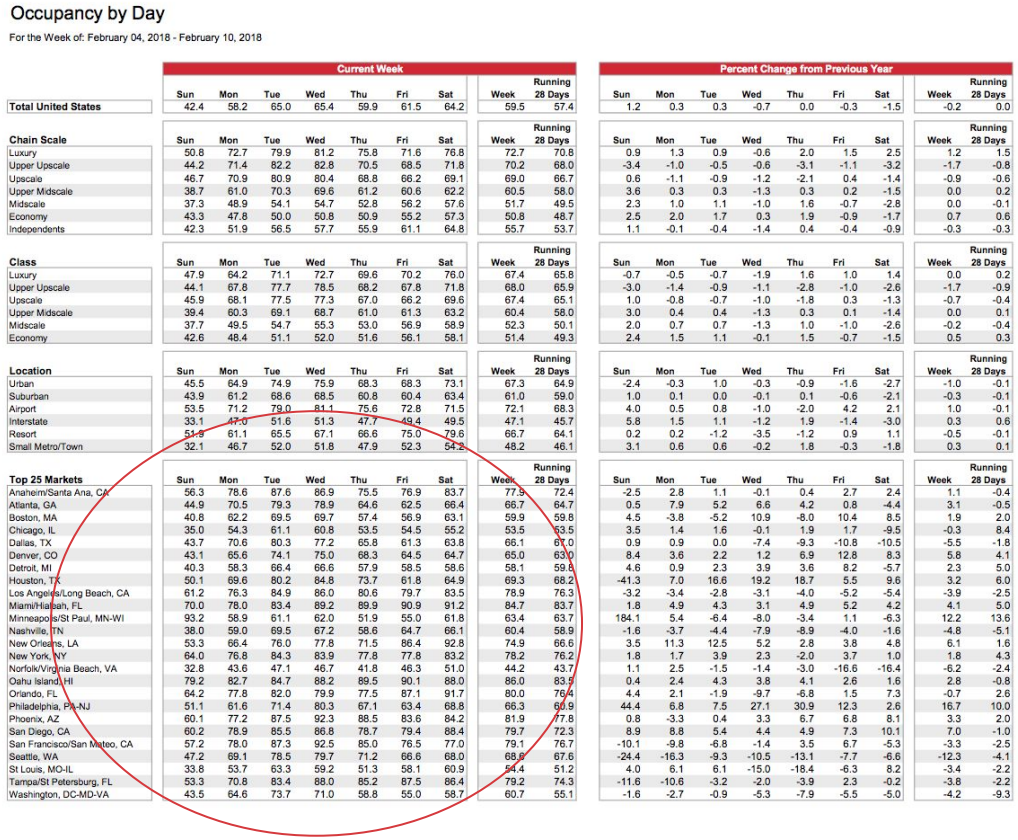
Shiny Dashboards



Leaflet



PDF Tools



Occupancy by Day

For the Week of: February 04, 2018 - February 10, 2018

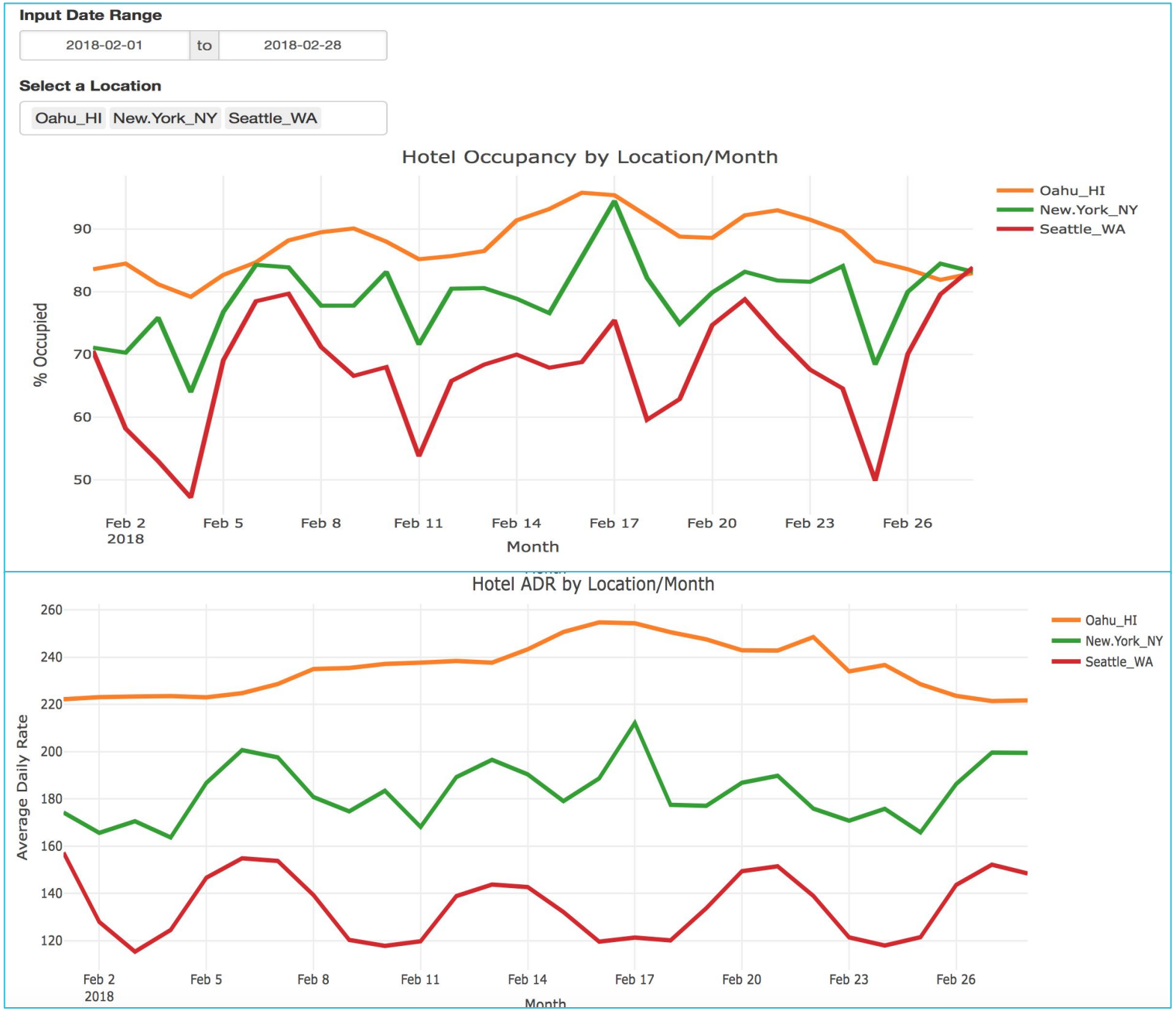
	Current Week									Percent Change from Previous Year								
	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Week	Running 28 Days	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Week	Running 28 Days
Total United States	42.4	58.2	65.0	65.4	59.9	61.5	64.2	59.5	57.4	1.2	0.3	0.3	-0.7	0.0	-0.3	-1.5	-0.2	0.0
Chain Scale																		
Luxury	50.8	72.7	79.9	81.2	75.8	71.6	76.8	72.7	70.8	0.9	1.3	0.9	-0.6	2.0	1.5	2.5	1.2	1.5
Upper Upscale	44.2	71.4	82.2	82.8	70.5	68.5	71.8	70.2	68.0	-3.4	-1.0	-0.5	-0.6	-3.1	-1.1	-3.2	-1.7	-0.8
Upscale	46.7	70.9	80.9	80.4	68.8	66.2	69.1	69.0	66.7	0.6	-1.1	-0.9	-1.2	-2.1	0.4	-1.4	-0.9	-0.6
Upper Midscale	38.7	61.0	70.3	69.6	61.2	60.6	62.2	60.5	58.0	3.6	0.3	0.3	-1.3	0.3	0.2	-1.5	0.0	0.2
Midscale	37.3	48.9	54.1	54.7	52.8	56.2	57.6	51.7	49.5	2.3	1.0	1.1	-1.0	1.6	-0.7	-2.8	0.0	-0.1
Economy	43.3	47.8	50.0	50.8	50.9	55.2	57.3	50.8	48.7	2.5	2.0	1.7	0.3	1.9	-0.9	-1.7	0.7	0.6
Independents	42.3	51.9	56.5	57.7	55.9	61.1	64.8	55.7	53.7	1.1	-0.1	-0.4	-1.4	0.4	-0.4	-0.9	-0.3	-0.3
Class																		
Luxury	47.9	64.2	71.1	72.7	69.6	70.2	76.0	67.4	65.8	-0.7	-0.5	-0.7	-1.9	1.6	1.0	1.4	0.0	0.2
Upper Upscale	44.1	67.8	77.7	78.5	68.2	67.8	71.8	68.0	65.9	-3.0	-1.4	-0.9	-1.1	-2.8	-1.0	-2.6	-1.7	-0.9
Upscale	45.9	68.1	77.5	77.3	67.0	66.2	69.6	67.4	65.1	1.0	-0.8	-0.7	-1.0	-1.8	0.3	-1.3	-0.7	-0.4
Upper Midscale	39.4	60.3	69.1	68.7	61.0	61.3	63.2	60.4	58.0	3.0	0.4	0.4	-1.3	0.3	0.1	-1.4	0.0	0.1
Midscale	37.7	49.5	54.7	55.3	53.0	56.9	58.9	52.3	50.1	2.0	0.7	0.7	-1.3	1.0	-1.0	-2.6	-0.2	-0.4
Economy	42.6	48.4	51.1	52.0	51.6	56.1	58.1	51.4	49.3	2.4	1.5	1.1	-0.1	1.5	-0.7	-1.5	0.5	0.3
Location																		
Urban	45.5	64.9	74.9	75.9	68.3	68.3	73.1	67.3	64.9	-2.4	-0.3	1.0	-0.3	-0.9	-1.6	-2.7	-1.0	-0.1
Suburban	43.9	61.2	68.6	68.5	60.8	60.4	63.4	61.0	59.0	1.0	0.1	0.0	-0.1	0.1	-0.6	-2.1	-0.3	-0.1
Airport	53.5	71.2	79.0	81.1	75.6	72.8	71.5	72.1	68.3	4.0	0.5	0.8	-1.0	-2.0	4.2	2.1	1.0	-0.1
Interstate	33.1	47.0	51.6	51.3	47.7	49.4	49.5	47.1	45.7	5.8	1.5	1.1	-1.2	1.9	-1.4	-3.0	0.3	0.6
Resort	51.9	61.1	65.5	67.1	66.6	75.0	79.6	66.7	64.1	0.2	0.2	-1.2	-3.5	-1.2	0.9	1.1	-0.5	-0.1
Small Metro/Town	32.1	46.7	52.0	51.8	47.9	52.3	54.2	48.2	46.1	3.1	0.6	0.6	-0.2	1.8	-0.3	-1.8	0.3	0.1
Top 25 Markets																		
Anaheim/Santa Ana, CA	56.3	78.6	87.6	86.9	75.5	76.9	83.7	77.9	72.4	-2.5	2.8	1.1	-0.1	0.4	2.7	2.4	1.1	-0.4
Atlanta, GA	44.9	70.5	79.3	78.9	64.6	62.5	66.4	66.7	64.7	0.5	7.9	5.2	6.6	4.2	0.8	-4.4	3.1	-0.5
Boston, MA	40.8	62.2	69.5	69.7	57.4	56.9	63.1	59.9	59.8	4.5	-3.8	-5.2	10.9	-8.0	10.4	8.5	1.9	2.0
Chicago, IL	35.0	54.3	61.1	60.8	53.5	54.5	55.2	53.5	53.5	3.5	1.4	1.6	-0.1	1.9	1.7	-9.5	-0.3	8.4
Dallas, TX	43.7	70.6	80.3	77.2	65.8	61.3	63.8	66.1	67.0	0.9	0.9	0.0	-7.4	-9.3	-10.8	-10.5	-5.5	-1.8
Denver, CO	43.1	65.6	74.1	75.0	68.3	64.5	64.7	65.0	63.0	8.4	3.6	2.2	1.2	6.9	12.8	8.3	5.8	4.1
Detroit, MI	40.3	58.3	66.4	66.6	57.9	58.5	58.6	58.1	59.8	4.6	0.9	2.3	3.9	3.6	8.2	-5.7	2.3	5.0
Houston, TX	50.1	69.6	80.2	84.8	73.7	61.8	64.9	69.3	68.2	-41.3	7.0	16.6	19.2	18.7	5.5	9.6	3.2	6.0
Los Angeles/Long Beach, CA	61.2	76.3	84.9	86.0	80.6	79.7	83.5	78.9	76.3	-3.2	-3.4	-2.8	-3.1	-4.0	-5.2	-5.4	-3.9	-2.5
Miami/Hialeah, FL	70.0	78.0	83.4	89.2	89.9	90.9	91.2	84.7	83.7	1.8	4.9	4.3	3.1	4.9	5.2	4.2	4.1	5.0
Minneapolis/St Paul, MN-WI	93.2	58.9	61.1	62.0	51.9	55.0	61.8	63.4	63.7	184.1	5.4	-6.4	-8.0	-3.4	1.1	-6.3	12.2	13.6
Nashville, TN	38.0	59.0	69.5	67.2	58.6	64.7	66.1	60.4	58.9	-1.6	-3.7	-4.4	-7.9	-8.9	-4.0	-1.6	-4.8	-5.1
New Orleans, LA	53.3	66.4	76.0	77.8	71.5	86.4	92.8	74.9	66.6	3.5	11.3	12.5	5.2	2.8	3.8	4.8	6.1	1.6
New York, NY	64.0	76.8	84.3	83.9	77.8	77.8	83.2	78.2	76.2	1.8	1.7	3.9	2.3	-2.0	3.7	1.0	1.8	4.3
Norfolk/Virginia Beach, VA	32.8	43.6	47.1	46.7	41.8	46.3	51.0	44.2	43.7	1.1	2.5	-1.5	-1.4	-3.0	-16.6	-16.4	-6.2	-2.4
Oahu Island, HI	79.2	82.7	84.7	88.2	89.5	90.1	88.0	86.0	83.9	0.4	2.4	4.3	3.8	4.1	2.6	1.6	2.8	-0.8
Orlando, FL	64.2	77.8	82.0	79.9	77.5	87.1	91.7	80.0	76.4	4.4	2.1	-1.9	-9.7	-6.8	1.5	7.3	-0.7	2.6
Philadelphia, PA-NJ	51.1	61.6	71.4	80.3	67.1	63.4	68.8	66.3	60.9	44.4	6.8	7.5	27.1	30.9	12.3	2.6	16.7	10.0
Phoenix, AZ	60.1	77.2	87.5	92.3	88.5	83.6	84.2	81.9	77.8	0.8	-3.3	0.4	3.3	6.7	6.8	8.1	3.3	2.0
San Diego, CA	60.2	78.9	85.5	86.8	78.7	79.4	88.4	79.7	72.3	8.9	8.8	5.4	4.4	4.9	7.3	10.1	7.0	-1.0
San Francisco/San Mateo, CA	57.2	78.0	87.3	92.5	85.0	76.5	77.0	79.1	76.7	-10.1	-9.8	-6.8	-1.4	3.5	6.7	-5.3	-3.3	-2.5
Seattle, WA	47.2	69.1	78.5	79.7	71.2	66.6	68.0	68.6	67.6	-24.4	-16.3	-9.3	-10.5	-13.1	-7.7	-6.6	-12.3	-4.1
St Louis, MO-IL	33.8	53.7	63.3	59.2	51.3	58.1	60.9	54.4	51.2	4.0	6.1	6.1	-15.0	-18.4	-6.3	8.2	-3.4	-2.2
Tampa/St Petersburg, FL	53.3	70.8	83.4	88.0	85.2	87.5	86.4	79.2	74.3	-11.6	-10.6	-3.2	-2.0	-3.9	2.3	-0.2	-3.8	-2.2
Washington, DC-MD-VA	43.5	64.6	73.7	71.0	58.8	55.0	58.7	60.7	55.1	-1.6	-2.7	-0.9	-5.3	-7.9	-5.5	-5.0	-4.2	-9.3

```
#getting page 7 from the file
get_file_from_date <- function(start_date) {
  filename <- paste(start_date, ".pdf", sep = "")
  file_text <- pdf_text(filename)
```

+ a lot more code & magic

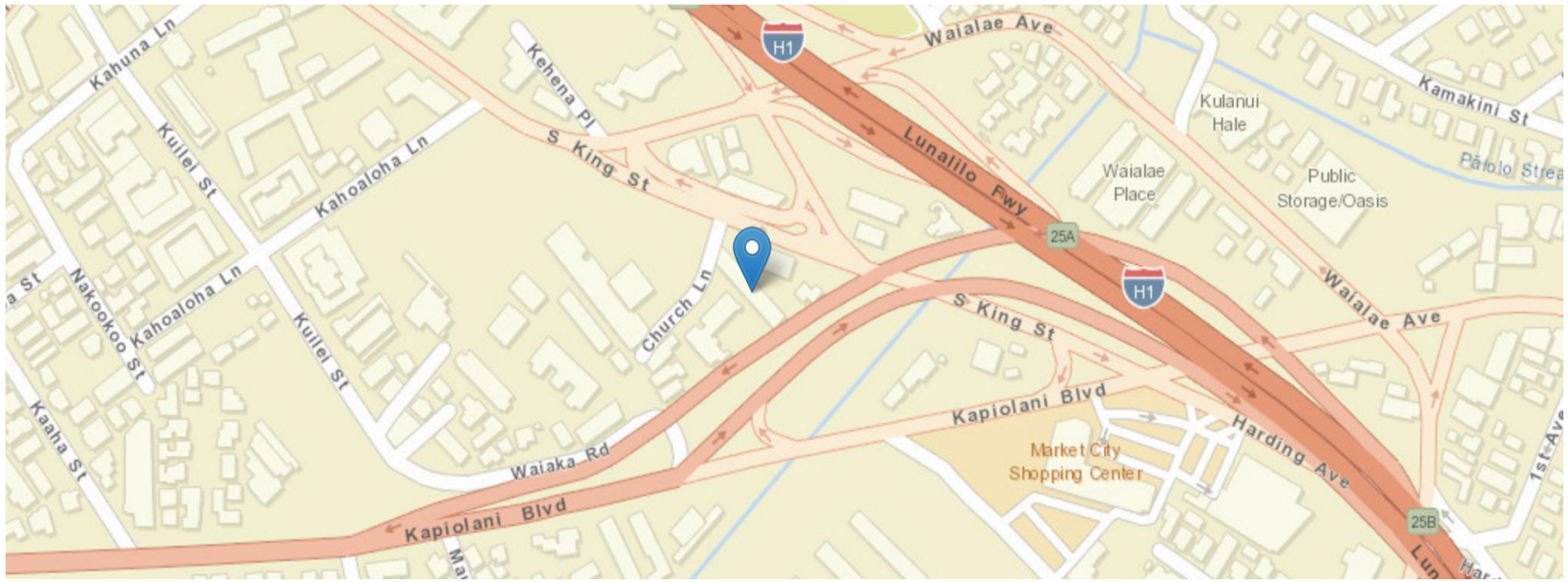
Date	DayofWeek	Anaheim/SantaAna_CA	Atlanta_GA	Boston_MA	Chicago_IL	Dallas_TX	Denver_CO	Detroit_MI
2018-02-04	Sunday	56.3	44.9	40.8	35.0	43.7	43.1	40.3
2018-02-05	Monday	78.6	70.5	62.2	54.3	70.6	65.6	58.3
2018-02-06	Tuesday	87.6	79.3	69.5	61.1	80.3	74.1	66.4
2018-02-07	Wednesday	86.9	78.9	69.7	60.8	77.2	75.0	66.6
2018-02-08	Thursday	75.5	64.6	57.4	53.5	65.8	68.3	57.9
2018-02-09	Friday	76.9	62.5	56.9	54.5	61.3	64.5	58.5
2018-02-10	Saturday	83.7	66.4	63.1	55.2	63.8	64.7	58.6

Shiny Dashboards

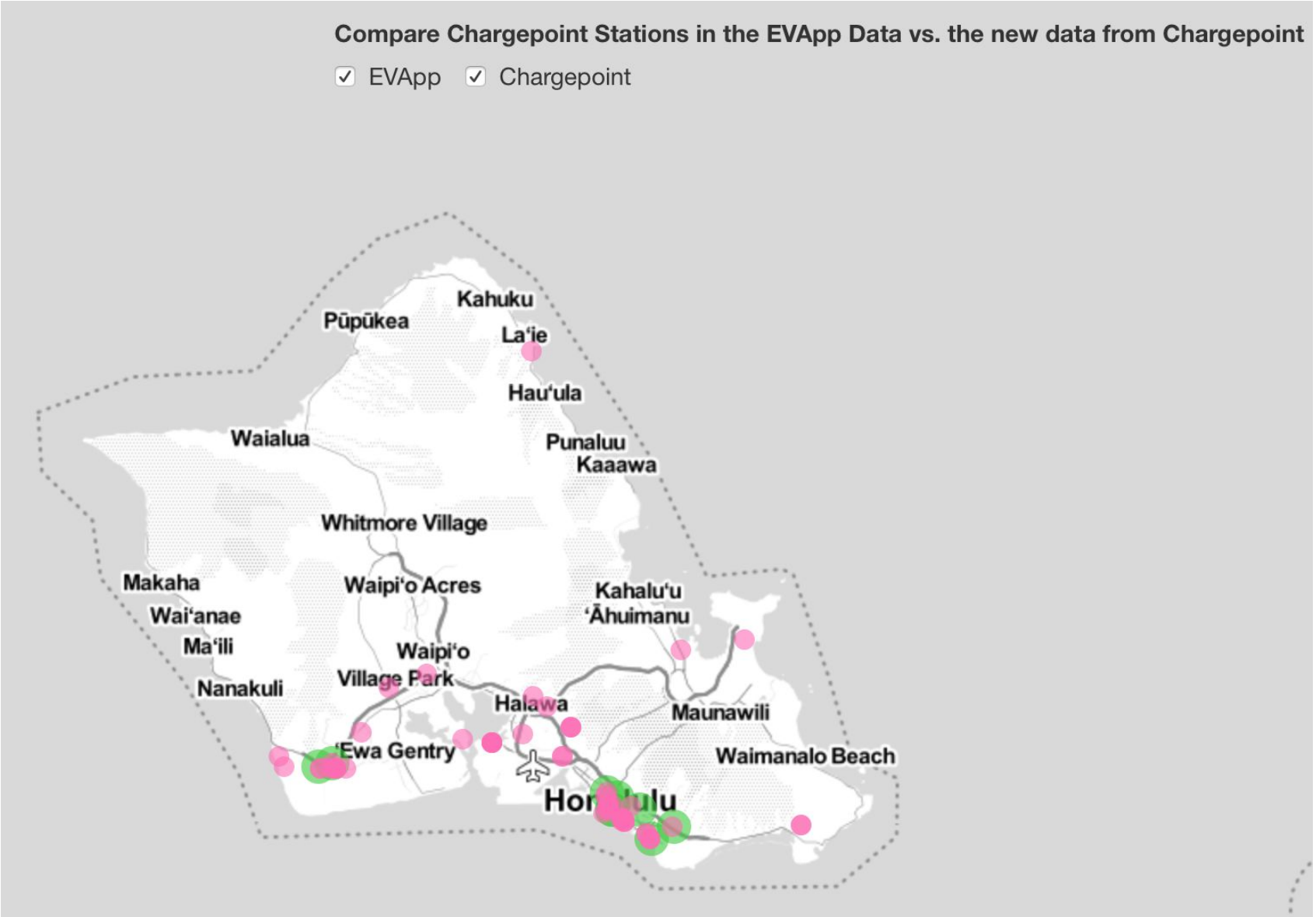


Leaflet

```
1  #install.packages("leaflet")
2  library(leaflet)
3  library(RColorBrewer)
4  library(shiny)
5
6  m <- leaflet()
7  m <- addTiles(m)
8  m <- addMarkers(m, lng = -157.817508, lat = 21.289207)
9
10 m %>% addProviderTiles(providers$Stamen.Watercolor)
```



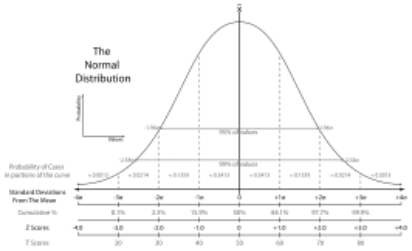
EV Charging App



The Future

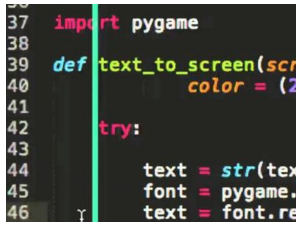
What's Next?

Runjini



Statistics Review

DataCamp



Python?

Practice!!

Jon

Kaggle Competitions

Stock Valuation

Tori

Shiny

Leaflet

DataCamp

Google those error messages


WE DID IT
TIME TO CELEBRATE
memegenerator.net

The end.

Mahalo!