Twee-Shirts

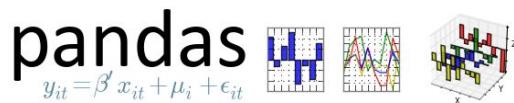*Gathering Live Hashtags from Sporting Events*

by Jason Hortsch

Business Motivation

- Screen print t-shirts directly outside an arena after a sports game

- The idea is to capture trends that occurred during a specific game
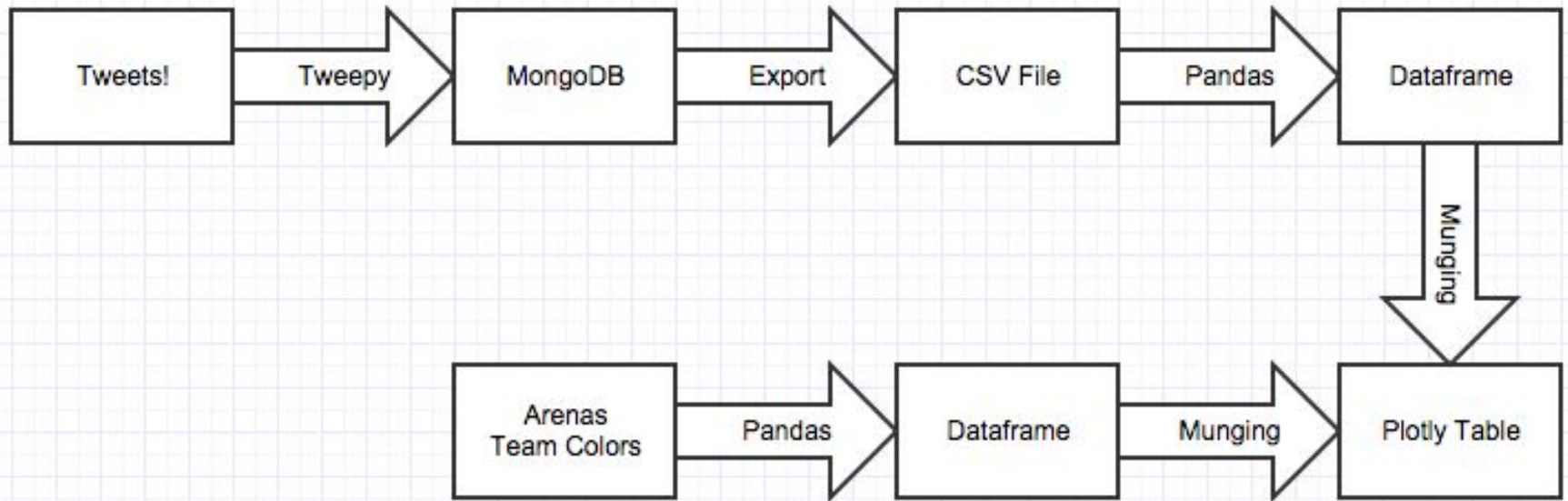
- Portable screen printer:

System Rundown

1. Stream in tweets using Tweepy

2. Store tweets in a Mongo database

3. Export tweets to a csv file

4. Import the file into a Pandas dataframe

5. Match team color and arena with team

6. Manipulate dataframe to find most popular tweets

7. Print color-coded table of most popular tweets with Plotly

Architectural Drawing

Picture of Tweets Coming In

Example Output

Team	Hashtag	Count	Rank
Celtics	#Celtics	1742	1
Celtics	#PORatBOS	563	2
Celtics	#NBA	421	3
Celtics	#Blazers	191	4
Celtics	#celtics	188	5
Celtics	#CelticsWin.	158	6
Celtics	#GreenRunsDeep	113	7
Celtics	#SCThrowback	109	8
Celtics	#CelticsNation	95	9
Celtics	#Blazzers	87	10
Celtics	#trailblazers	83	11
Celtics	#CelticsWin!!	80	12
Celtics	#Boston	80	13
Celtics	#Trailblazers	76	14

Example Output 2

Team	Hashtag	Count	Rank
Raptors	#Raptors	1296	1
Raptors	#RTZ	1285	2
Raptors	#WeTheNorth	700	3
Raptors	#NBALeaguePass.	313	4
Raptors	#LeaguePassAlert	204	5
Raptors	#PORatTOR	204	6
Raptors	#NBA	149	7
Raptors	#Blazers	147	8
Raptors	#raptors	133	9
Raptors	#rtz	98	10
Raptors	#Toronto	73	11
Raptors	#TrailBlazers	66	12
Raptors	#...	58	13
Raptors	#wethenorth	38	14

Mockup:

Lessons Learned

- Use pre-existing resources

- Read documentation

- Try / excepts are your friend

- Do things in small pieces

- Twitter data not as big as I thought

Version 2.0

- Make things even more streamlined

- Remove the manual components

- Do all games automatically (using an NBA or ESPN api?)

- After Mongo, put tweets in a RDBMS and query with SQL

- Create a web app?

- Make my dashboard more 'live'