

# PGA Tour and Weather Data

---

PGA ShotLink Data Team

# Background

- The PGA Tour offers shot level data through the “ShotLink® Intelligence” program
- Weather could have an effect on player performance
- Quality of weather data provided with ShotLink is lacking
  - AM/PM wind speed, direction
- DarkSky API provides hourly, historical weather data
  - Both qualitative and quantitative measures
    - Type of precipitation
    - Wind speed/direction
    - Temperature

**I DONT THINK THE HEAVY  
STUFF IS GONNA COME DOWN**



**FOR QUITE A WHILE**

# Goal(s)

- Provide python package to help data scientists manage ShotLink and DarkSky
  - Simplify the process to retrieve corresponding weather data
- Provide example analyses leveraging shot level data with hourly weather data

# Data Used

- PGA ShotLink

- **Overview**

- Collection of shot level and aggregate data from PGA Tour events
    - Focus on Strokes Gained metric
      - How well did a given shot perform based on how the average PGA Tour player plays

- **Process**

- Downloaded semicolon delimited files from ShotLink portal

- **Limitations**

- Strokes Gained based on 5 year tour average
    - Access to data restricted by application

- DarkSky API

- **Overview**

- Provides an API that allows users to query hourly historical weather data

- **Process**

- Used [darkskylib](#) python package to make calls to the API

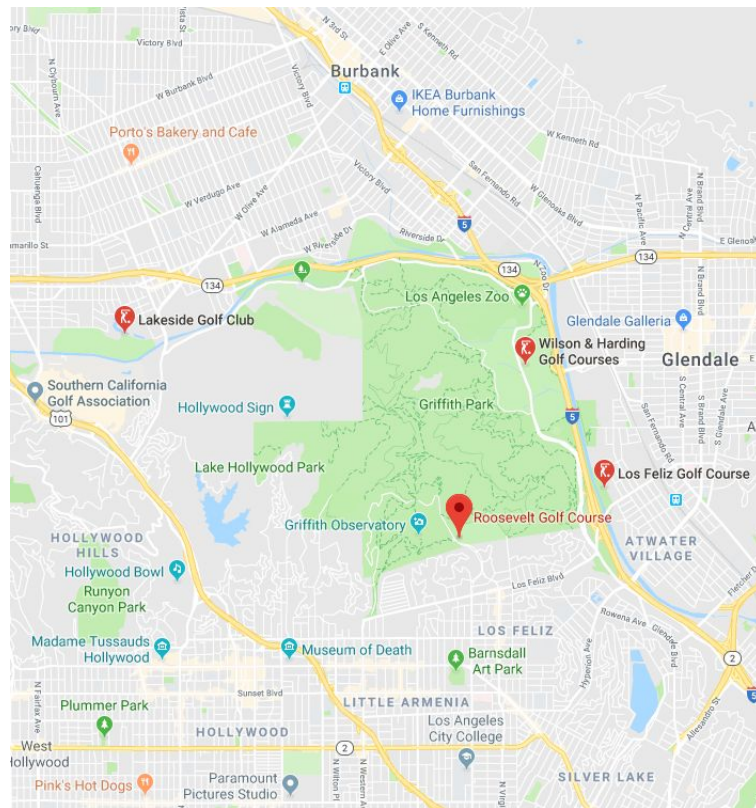
- **Limitations**

- Hour is the finest granularity provided
    - Needs latitude and longitude to retrieve data
    - 1,000 free calls per day
      - \$1 per 10,000 calls after that



# Data Used

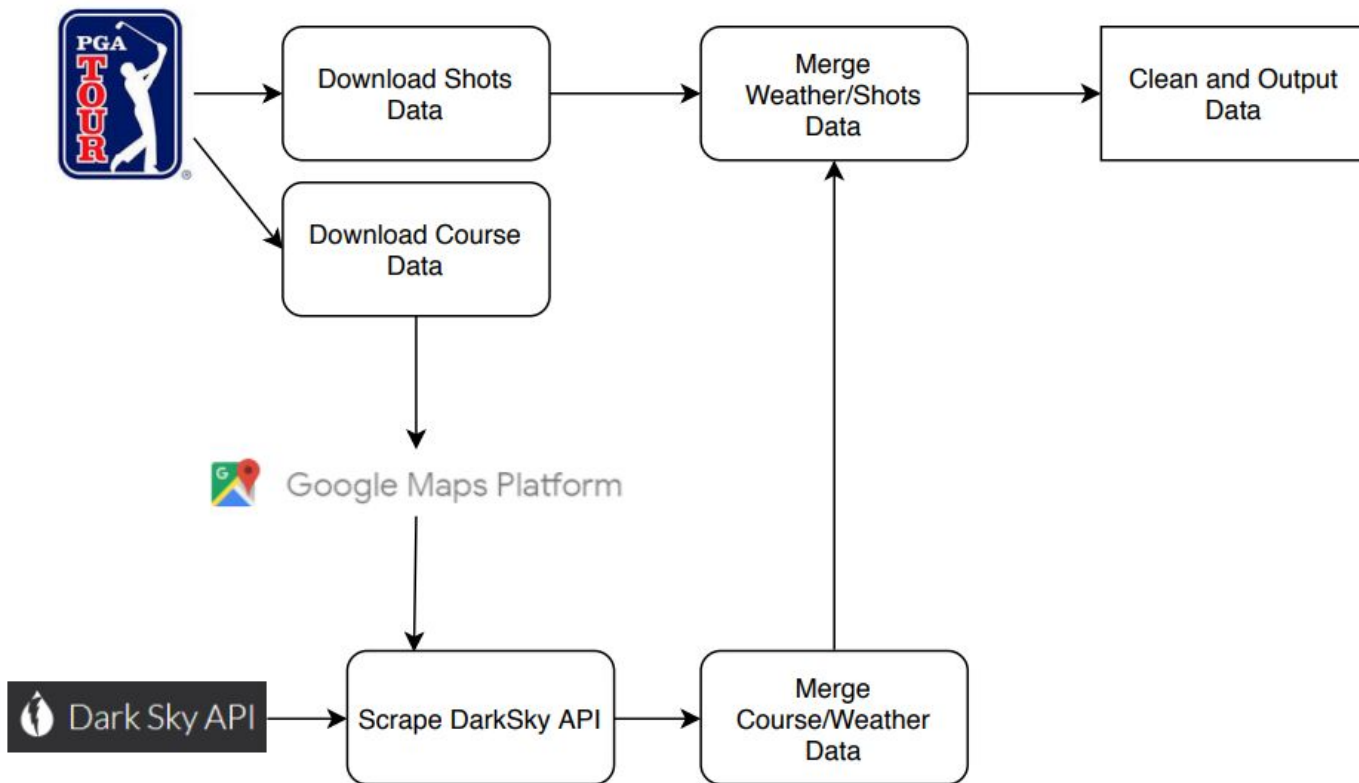
- Google Geocoding API
  - **Overview**
    - Maps a given address/location name to latitude and longitude
  - **Process**
    - Python script to iterate over unique course names in ShotLink golf course data set
  - **Limitations**
    - Only have course name from ShotLink data set



# Use Cases

- Overall Goal
  - Data scientists can analyze PGA ShotLink data and weather data, together.
- Example Research Questions
  - How does the presence of rain/wind/cloud cover affect golfer performance?
  - How does intensity of rain/wind/cloud cover affect golfer performance?
  - Which golfers are most affected by weather events?

# Design Spec





# Github Repo

Project Repository:

- [https://github.com/awfuldynne/golf\\_course\\_project](https://github.com/awfuldynne/golf_course_project)

# Demo



# Lessons Learned

- Many gotchas to look out for when working with date/time data
- A simple Linear Regression model is probably not sufficient to deduce how weather conditions affect shot performance
- The overwhelming majority of our shot events took place during nice weather. This makes it difficult to find correlations with the more 'extreme' weather conditions.
- The free version of Travis CI (public repo) does work with a private repo dependency, using Git submodules and bespoke encryption/decryption

# Future Work

- Further investigation into code coverage solutions
  - [Coveralls](#) doesn't always update with coverage data from Travis
  - But we implemented [CodeCov](#) and it seems better
- Player-level analyses
  - Are certain players more effective in rain?
- Deduce wind direction relative to shot direction using wind bearing and sequential shot coordinate values
- Downsample shots in 'nice' weather
- Unit test all higher-level ShotLink data cleaning/merging scripts - right now we test all of the core functionality, but only some of the higher-level scripts that call the core functionality



Questions?