Welcome!

The following assessment test focuses on 3 areas:

- ETL
- API
- Data Modelling & Functions

Problems are taken from Housecall Pro (HCP) examples.

The test is limited to 3 hours. Please deliver **executable** code, a SQLite Dump, and answers (in as much detail as you're comfortable covering) to dataeng@housecallpro.com

Please share your thoughts as much as possible. It's not always about the right answer, but the way you tackle the problem.

1. Weather ETL

Housecall Pro values understanding weather seasonality for organizations that perform seasonal work. (Ie more air conditioners /heating vents are typically installed during hotter months) To give the business insight into seasonal weather patterns, Analytics would like a database table that provides weather pattern data for specific cities in the US.

Here is a list of specific cities we're interested in weather info for.

Given a list of city names, use the Yahoo Weather API (We recognize it's going to be deprecated soon. You can still make calls against it) and either Python or Scala to fetch data for those cities given the following constraints:

1. For each location and date, write the following to a single SQLite table
    - data_load_date
    - City
    - State
    - City Latitude
    - City Longitude
    - Forecast_date // the api returns a 10 day forecast from a given date
    - Low Temp
    - High Temp
    - Wind Speed

- Wind Chill
- Humidity

2. Next, please create another SQLite table with the following fields:
Date | City | State | Avg High Temp in Next 5 Days | Avg Wind Speed

3. If this were an ongoing ETL, what would be your plan for ensuring it ran reliably and without issues in production?

4. What processes would you want in place to ensure business expectations are appropriately set should anything go wrong with this ETL?

2. SQL

Consider the following SAAS_history table, containing SAAS subscription transactions. Note how there is one unique record for every day a plan changed for a given company. In other words, Company **1** didn't have any records loaded on **Jan 2** because it didn't have a plan change take place.

SAAS_history:

company_id	**plan_started_on**	**plan**
1	2017-01-01	free trial
1	2017-01-03	advanced
2	2017-02-14	free trial
2	2017-02-28	free
..	..	..

Please generate a SQL script that creates records defining what plan a company is on for any given day between Jan 1, 2017 and Dec 31, 2017.

Output:

company_id	**date**	**plan**
1	2017-01-01	free trial
1	2017-01-02	free trial

1	2017-01-03	advanced
1	2017-01-04	advanced
1	...	...
2	2017-02-14	free trial
2	2017-02-15	free trial
2	...	...
2	2017-02-28	free
2	2017-02-29	free
2	...	...

Feel free to make any assumptions you like about the database in which this hypothetical data lives.

1. Please share your SQL script with us.
2. Please communicate how/if your underlying database assumptions impacted the nature of your SQL code.

3. Data Modeling and Analysis

Using the data you loaded into the SQLite table from the Yahoo weather resource, we would like you to develop a function to query regional weather. Specifically, your function should take two arguments: **1** City name, and **2** Radial distance in miles. With that data, please generate a query programmatically, using either Python or Scala, to compute the following:

- How many cities fall within the given distance from the given city?
- What is the average humidity index for all those cities?
- What is the average temperature for all those cities?

Your function can simply return those answers as a dictionary to the console.

```
{
    “given_radius_in_miles”:  100,
    “origin_city”  : ‘Kansas City”,
    "num_cities_in_radius": 10,
    "mean_humidity":  90,
    “mean_temp” :  75
```

}

1. Please share any code you wrote to create this.
2. How could this script be improved/changed to make it production worthy?
3. How could this function be modified to take arguments as GET requests over HTTP?
    a. What challenges would you expect that you'd want to account for?
    b. Please communicate any assumptions you made

Please share code you wrote or sqlite database you generated to dataeng@housecallpro.com