Problem Set J Submission Form

Overview

Your Name	Hendi Kushta
Your SU Email	hkushta@syr.edu

Instructions

Put your name and SU email at the top. Answer these questions all from the lab. When asked to include screenshots, please follow the screen shot guidelines from the first homework.

Remember as you complete the homework it is not only about getting it right / correct. We will discuss the answers in class so it's important to articulate anything you would like to contribute to the discussion in your answer:

- If you feel the question is vague, include any assumptions you've made.
- If you feel the answer requires interpretation or justification provide it.
- If you do not know the answer to the question, articulate what you tried and how you are stuck.
- Highlight any doubts or questions you would like me to review.

This how you receive credit for answering questions which might not be correct. In addition, you must complete the reflection portion of the homework assignment for full credit. Since most answers will be similar this is an important part of your individual submission.

Complete Part II of this document first, then go back and complete the Reflection in Part I.

Part I - Reflection

Use this section to reflect on your learning. To achieve the highest grade on the assignment you must be as descriptive and personal as possible with your reflection.

1. As you completed this assignment, identify what you learned.

Query data in Elasticsearch using the REST API. Import and export data from Elasticserch into spark dataframes. Use Kibana to create index patterns and searchable fields. Build Dashboards, Maps and canvases in Kibana.

2. What barriers or challenges did you encounter while completing this assignment?

Question 1 is 1 challenge I faced, because the code, prints +4 hours from the localtime.

3. How prepared were you to complete this assignment? What can you do to be better prepared?

   I was prepared for the assignment. More exercises.

4. Rate your comfort level with this week's material. Use the rubric provided.

4 ==> I understand this material and can explain it to others.
3 ==> I understand this material.
2 ==> I somewhat understand the material but sometimes need guidance from others.
1 ==> I understand very little of this material and need extra help.

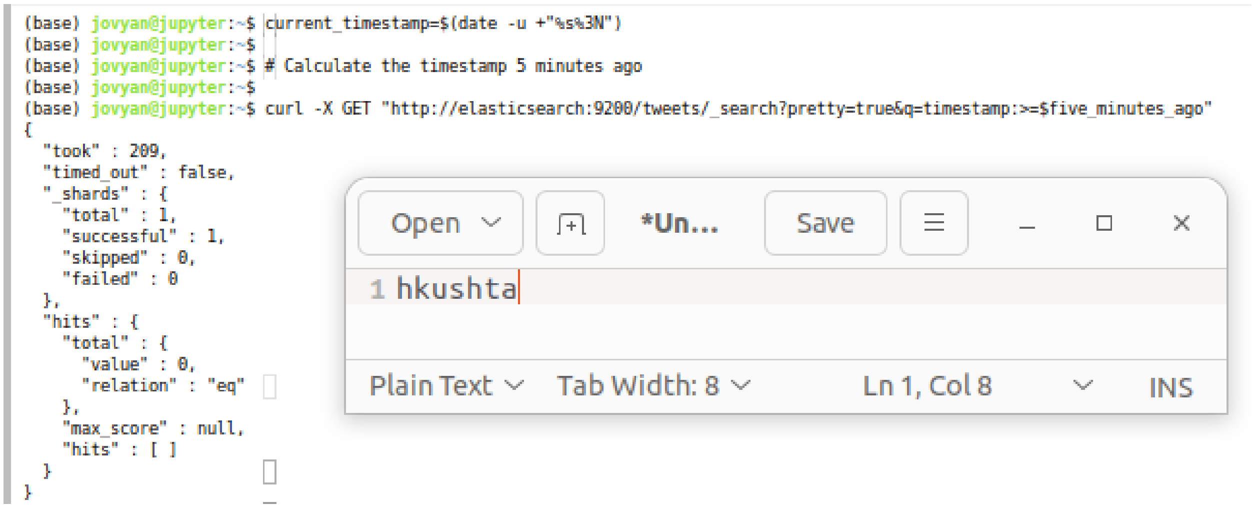
Part II – Questions

For each question, include a copy of the code required to complete the question along with a screenshot of the code and a screenshot of the output.

1. Turn on tweets from the example **J-Elasticsearch** notebook. Search for tweets in the last 5 minutes from the (windows or linux) with **curl** (do the math yourself – don't make this calculation automatic).

The issue with question 1 is that I do not know, why when we run the J-Search.ipynb, I get the timestamp_format in +4 hours than the current local time. I tried to change the python code in //content/simtweet/__init__.py to get the current time, but still I cannot.

So what I did in the beginning, I tried to get the current timestamp in milliseconds, then I have calculated the timestamp 5 minutes ago, and in the end I have tried to get the result through the search query.

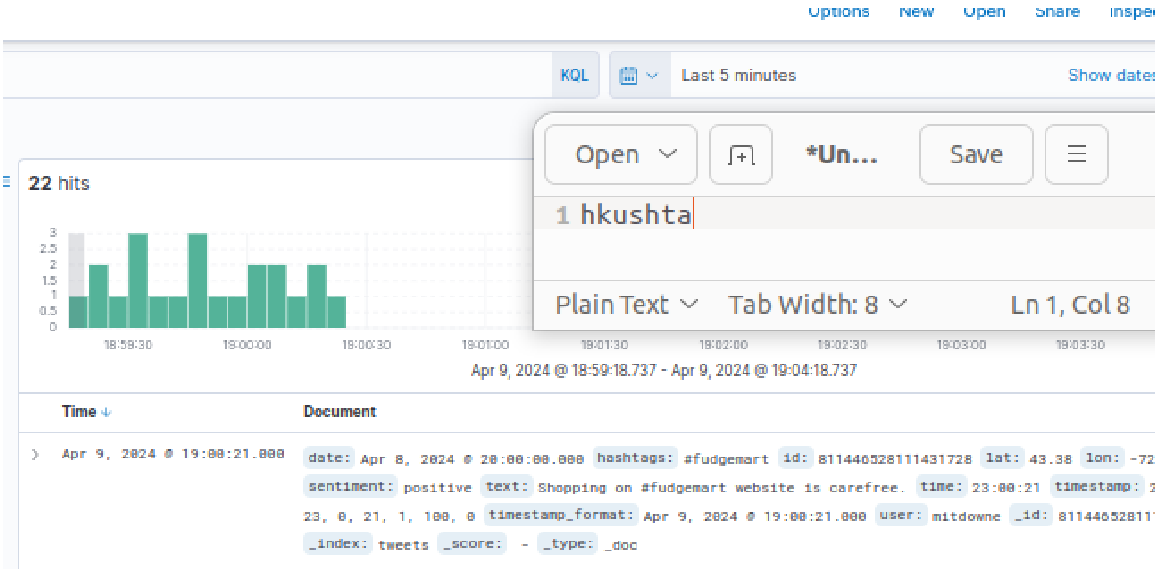


```
(base) jovyan@jupyter:~$ current_timestamp=$(date -u +"%s%3N")
(base) jovyan@jupyter:~$
(base) jovyan@jupyter:~$ # Calculate the timestamp 5 minutes ago
(base) jovyan@jupyter:~$
(base) jovyan@jupyter:~$ curl -X GET "http://elasticsearch:9200/tweets/_search?pretty=true&q=timestamp:>=$five_minutes_ago"
{
  "took" : 209,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 0,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  }
}
```

```
hkushta
```

2. Do the same thing from the Kibana UI for Elasticsearch. What must you do before you can do that and why? Turn off the tweets.

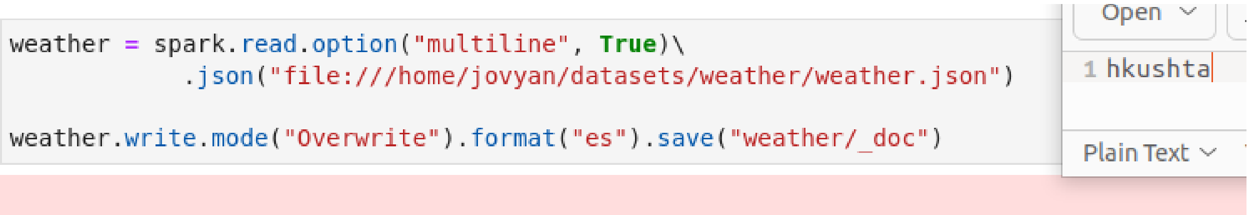
The first thing we have to do is to choose the timestamp_format or date format for the field of timestamp, because, we are trying to filter based on time.



Apr 9, 2024 @ 18:59:18.737 - Apr 9, 2024 @ 19:04:18.737

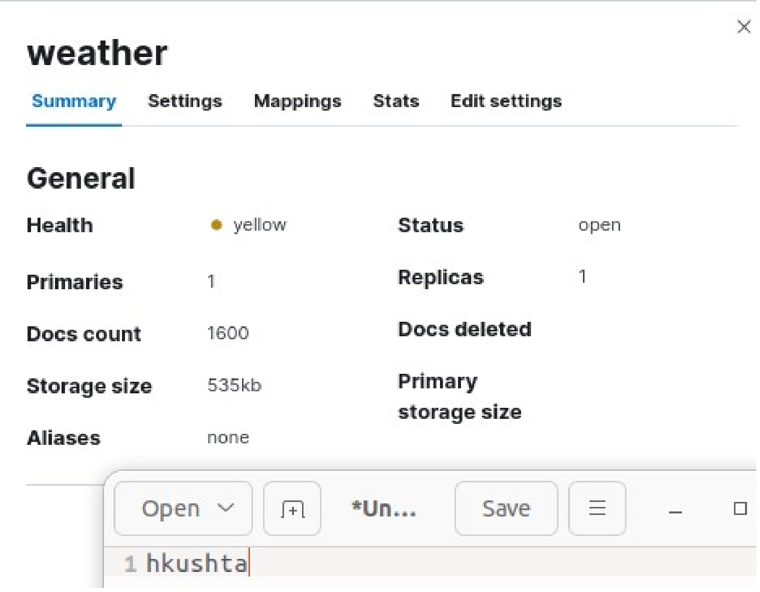
```
Time          Document
Apr 9, 2024 @ 19:00:21.000   date: Apr 8, 2024 @ 20:00:00.000 hashtags: #fudgemart id: 811446528111431728 lat: 43.38 lon: -72
                             sentiment: positive text: Shopping on #fudgemart website is carefree. time: 23:00:21 timestamp: 2
                             23, 0, 21, 1, 100, 0 timestamp_format: Apr 9, 2024 @ 19:00:21.000 user: mitdowne _id: 81144652811
                             _index: tweets _score: - _type: _doc
```

3. Write PySpark to load the 1,600 line weather data set into Elasticsearch under the index `weather` with default index type.



```
weather = spark.read.option("multiline", True)\
            .json("file:///home/jovyan/datasets/weather/weather.json")

weather.write.mode("Overwrite").format("es").save("weather/_doc")
```

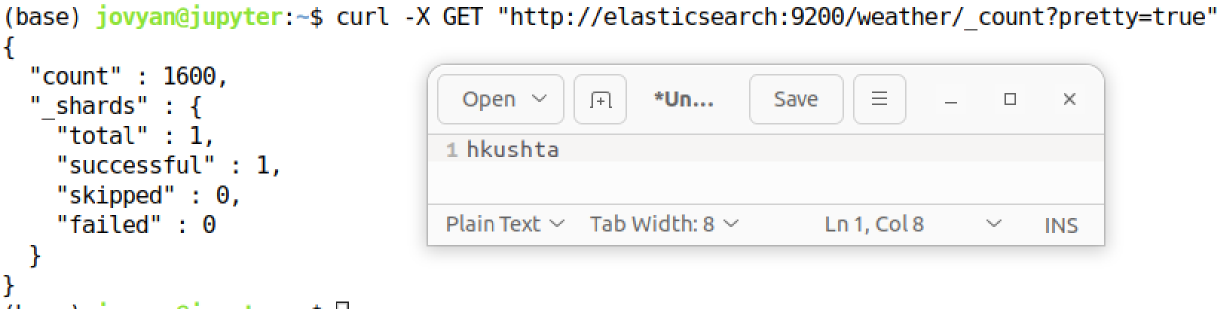


weather

General

Health	● yellow	Status	open
Primaries	1	Replicas	1
Docs count	1600	Docs deleted	
Storage size	535kb	Primary storage size	
Aliases	none		

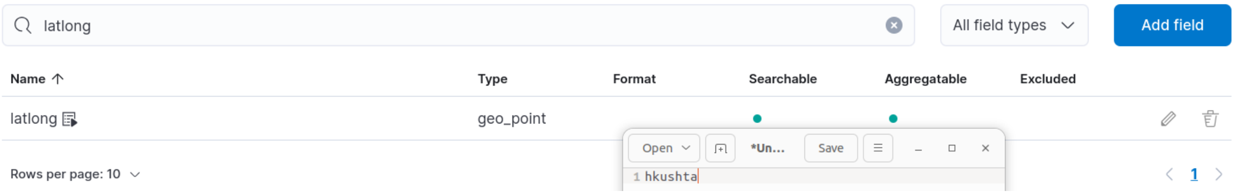
4. Use a **curl** command from the command line to hit the Elasticsearch API and demonstrate that there are 1,600 documents in the **weather** index.



```
(base) jovyan@jupyter:~$ curl -X GET "http://elasticsearch:9200/weather/_count?pretty=true"
{
  "count" : 1600,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  }
}
```

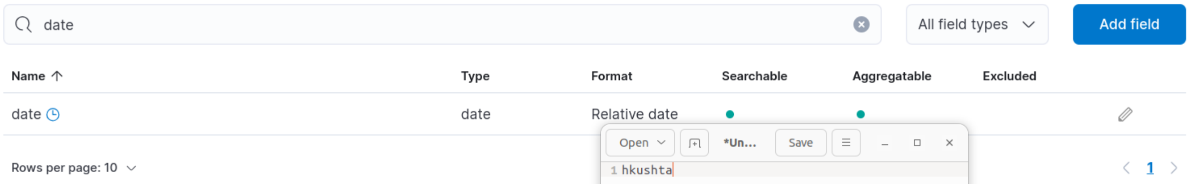
```
hkushta
```

5. Setup a **weather** index pattern in Kibana based on the **weather** index from Elasticsearch. Make sure you have a **geo_point** based on lat/lon type and have selected a **@timestamp** field using the date field. Provide a screenshot including the fields in question.



Name	Type	Format	Searchable	Aggregatable	Excluded
latlong	geo_point		●	●	

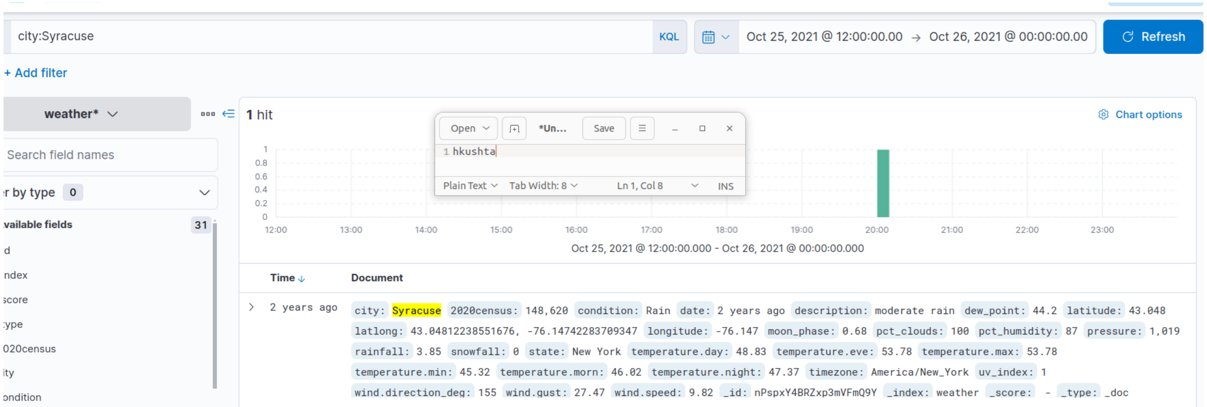
```
hkushta
```



Name	Type	Format	Searchable	Aggregatable	Excluded
date	date	Relative date	●	●	

```
hkushta
```

6. Demonstrate your Kibana index pattern is functional. For the most recent entry get the weather for a City of your choice.



city:Syracuse

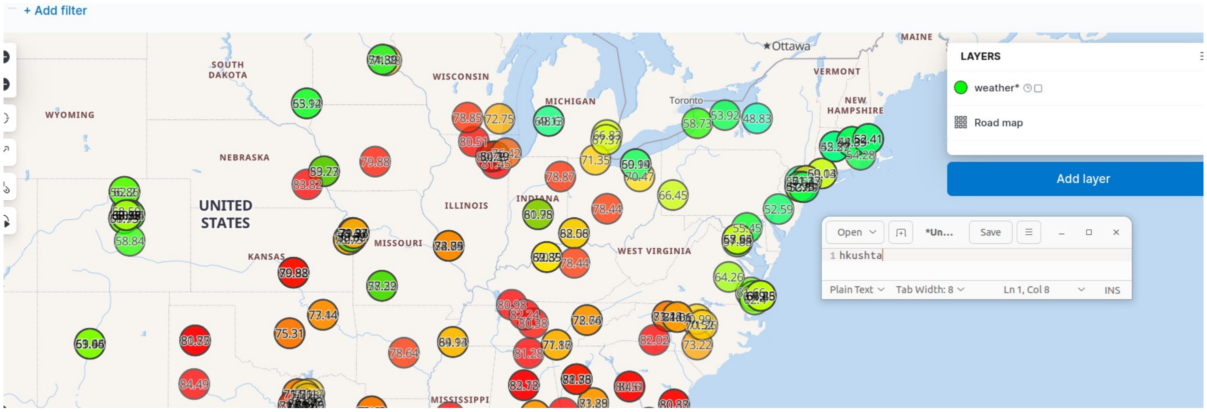
Oct 25, 2021 @ 12:00:00.00 → Oct 26, 2021 @ 00:00:00.00

2 years ago

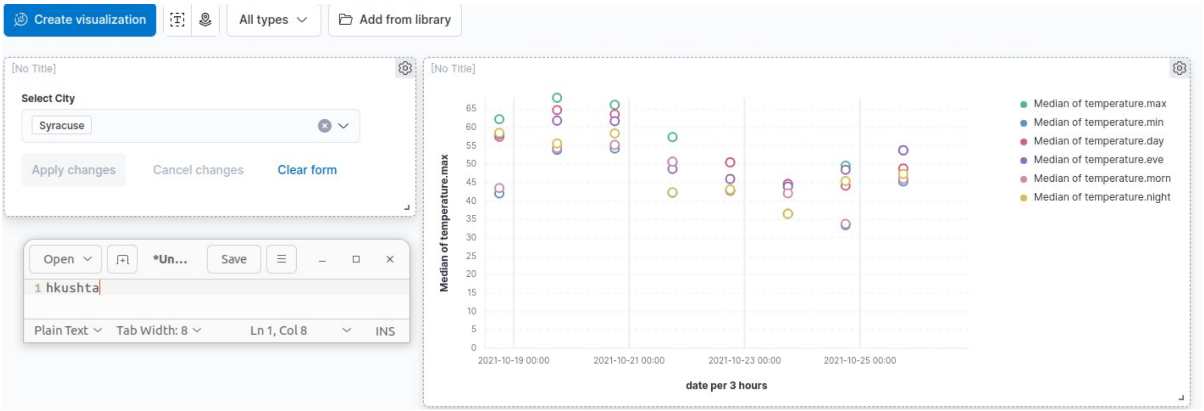
```
city: Syracuse 2020census: 148,620 condition: Rain date: 2 years ago description: moderate rain dew_point: 44.2 latitude: 43.048
latlong: 43.04812238551676, -76.14742283709347 longitude: -76.147 moon_phase: 0.68 pct_clouds: 100 pct_humidity: 87 pressure: 1,019
rainfall: 3.85 snowfall: 0 state: New York temperature.day: 48.83 temperature.eve: 53.78 temperature.max: 53.78
temperature.min: 45.32 temperature.morn: 46.02 temperature.night: 47.37 timezone: America/New_York uv_index: 1
wind.direction_deg: 155 wind.gust: 27.47 wind.speed: 9.82 _id: nPspxY4BRZxp3mVFmQ9Y _index: weather _score: - _type: _doc
```

7. Create a Kibana map displaying the weather locations for the most recent weather data. Use any layer(s) of data you wish. Provide a screenshot of the map with data points on it.

   The red circles are where temperatures are the highest, and as the circles change to blue color, the temperatures are the lowest.



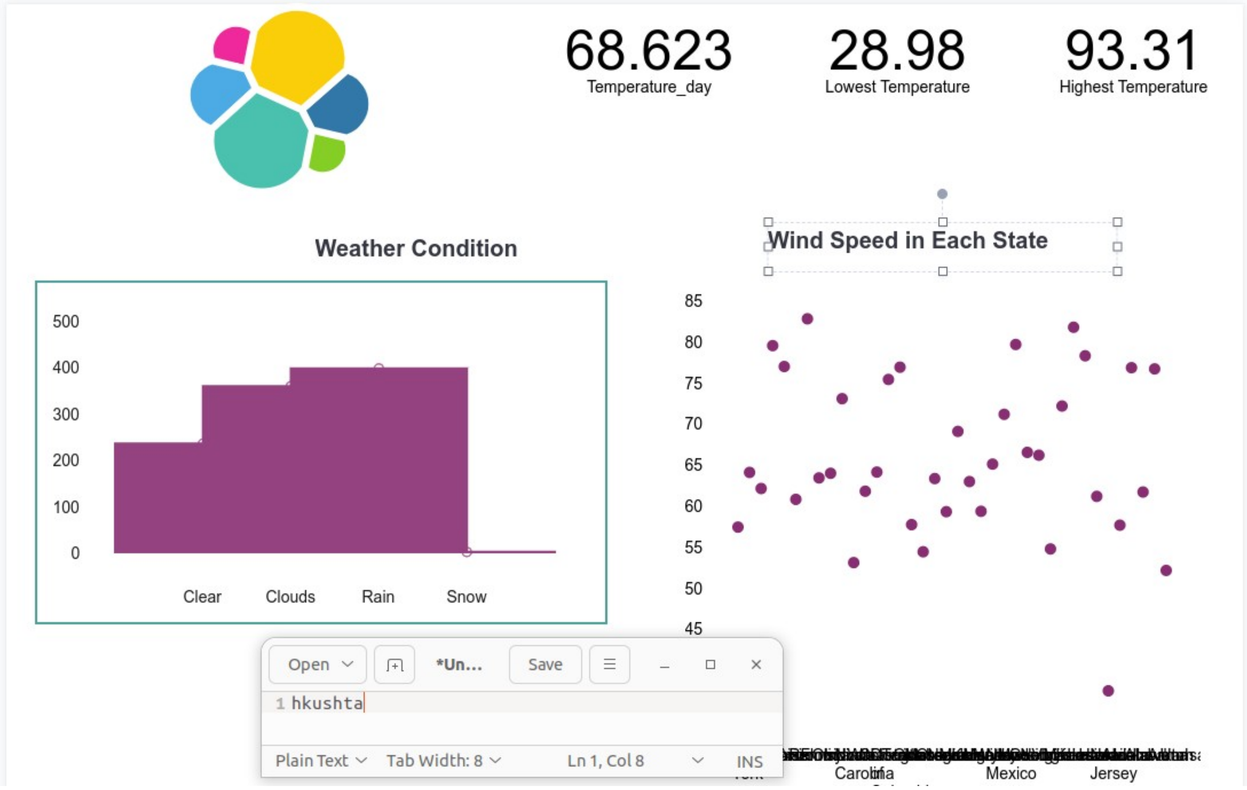
8. Create a Kibana dashboard which when you select a city, will display the average day time and nighttime temperature for that city, in addition to a line chart of the average daily high and lows for all data on that city. Provide a screenshot of the dashboard in action.



Select City: Syracuse

1 hkushta

9. Create a Kibana Canvas! Display at least 2 metrics and 2 charts. Decide which data you want to display and how you would like to present it. Provide a screenshot of the Canvas. You can do this for a set of cities or a single city.



68.623
Temperature_day

28.98
Lowest Temperature

93.31
Highest Temperature

1 hkushta