**CHALLENGE 6**

**REPORT**

**Introduction**:

The tourism industry greatly relies on the information provided online to potential clients, who have diverse interests in visiting sites all over the world. Tourists, on the other hand, benefit just as much from such information.

Information providers tailor vast troves of information on tourism from several sites to fit the interests of any particular type of traveler. This objective can be achieved using API (Application Programming Interface).

API facilitates communication between two applications. Data can be obtained from one application and implemented to create another. A string of unique combination of numbers and letters is generated by a website called an API key. The key is assigned to an ‘application-developer’ who may be interested in creating a special application based on the information available on the website.

**Analysis**:

In this Challenge, two websites as used:

https://[api.openweathermap.org](http://api.openweathermap.org/) , and <https://myprojects.geoapify.com> .

Part I

Part I is regarding weather data from various cities around the world.

A randomly generated combination of latitude and longitude, that happens to be in the vicinity of a city, can yield the name of the city. A list of over 500 cities is generated using this process.

Next, using the ‘weather’ website provided above, weather conditions at each location can be determined, visualized as graphs, and analyzed.

Since a randomly selected latitude-longitude pair may not always be near or in a city, the number of cities found is smaller than the number of combinations of latitudes and longitudes.

Weather data for each city is obtained from https://[api.openweathermap.org](http://api.openweathermap.org/).

Temperature vs. Latitude plot (Graph 1) of the cities illustrates that temperatures increase more steeply as one moves away form the South Pole toward the Equator. However, presently (today’s date in the title), the temperatures in the historically temperate regions (20° - 40°) are higher than in the equatorial belt!

This is due to the peak of summer in the northern hemisphere and the current season of winter in the southern hemisphere.

Latitudes > 40° show a gradual decline in temperature toward higher latitudes. The rate of decline in temperatures from the equator to the north pole is lower than the rate of rise in temperatures from the south pole to the equator. In the southern hemisphere, temperatures fall below 0°C in some cities between latitudes -40° and -60° and that there a no cities for latitudes < -60°. Whereas in the northern hemisphere, temperatures are above 0°C in all the cities found up to latitude 80°.

Graph 1

A graph of a diagram

Description automatically generated

Graph 2 is an illustration of humidity in the cities on the list. Humidity is measured by the amount of moisture in the air.

While humidity can be the measure of a season, it could also be a year-round feature of a location/region. For example, tropical locations in the equatorial region like Indonesia may have high humidity ( >80%), and landlocked regions in the same latitudinal range like Sub-Saharan Africa can have low humidity (< 10%).

The graph shows a number of cities with low humidity, but there are more cities with high humidity in the range of latitudes between -60° and 70°.

Graph 2

A diagram of a city

Description automatically generated

Graph 3 shows the relationship between latitude and wind speed for the cities. Most cities have wind speeds less than 10 m/s.

Comparing Graph 2 with Graph 3, a rough observation of opposite trends can be seen. Generally, though not always, high humidity implies low wind speeds.

Graph 3

A graph of blue dots

Description automatically generated

Graph 4 shows cloudiness vs. latitude. Cloudiness may not be as predictable as the other weather features, *viz*., temperature, humidity and wind speed.

However, some trends can be observed with the available data. Between latitudes 0° and 20°, a large number of cities show 100% cloudiness. This may be due to the monsoon season over the equatorial region, predominantly north of the equator. High percentage of cloudiness is also observed for cities between 40° and 70°. This may be due to the rain and storm season in the subtropics in the northern hemisphere.

Close to 0% cloudiness between latitudes 20° and 50° and between -10° and -40° are due the presence of arid regions, as in the middle east and in the southern portion of South America, respectively.

Graph 4

A graph of blue dots

Description automatically generated

Linear Regression of each weather feature by northern and southern hemisphere

Linear regression is a graphical statistical method, that assists in visualizing and determining trends that are given quantitative values. The values obtained can be studied further in predictability analysis of two variables.

In our context, the values of interest to us are slope, intercept and *r*-value.

Slope describes the rise and fall of a variable with respect to another variable. Based on the slope a straight line can be generated which would show a predictable trend between two features.

However, because the data in a practical situation doesn’t fall in a single straight line, different observers, may choose different straight lines, but with slopes falling in a narrow range of value.

Hence the introduction of *r*-value, also called the correlation coefficient. The slope of a regression line and *r*-value are close, but **not equal**. *r*-value gives us an assessment of linearity of a set of data.

Another point of difference: slope may have units, but *r*-value does not.

Just as slope, *r*-values range between –1 and +1, with 0 being ‘no linear relationship’ between two variables.

In the following graphs, S implies graph for Southern hemisphere (left) and N implies graph for Northern hemisphere (right).

Graph 5 data:

*r*(S) : +0.82 slope(S): +0.53

*r*(N) : -0.57 slope(N): -0.23

r(S) > *r*(N), implies that temperature vs. latitude for southern hemisphere is more linear than that in the northern hemisphere.

Slope(S) > slope(N), there is more variation in temperature with latitude in the ‘south’ than in the ‘north’.

The lower slope value for the northern hemisphere offers more choices for vacationers looking for warmer places to visit.

The intercepts indicate that values of temperature near the equator could range between 30°C and 35°C.

Graph 5 (S, N)

A comparison of a graph

Description automatically generated

Graph 6 data:

*r*(S) : -0.10 slope(S): -0.16

*r*(N) : +0.01 slope(N): +0.01

In the plots for humidity vs. latitude, it is difficult to determine, if there is any clear relationship between the two, both in the northern and southern hemispheres. The *r*-values and close to 0, and so are the slopes.

There are more data points away from the straight lines than in the neighborhood of the lines.

Slope(S) = 0.16, may give a false sense of linearity between the variables.

Thus, in the case of humidity vs. temperature, any conclusion regarding linearity will yield no useful result from a tourism perspective.

Graph 6 (S, N)

A screenshot of a graph

Description automatically generated

Graph 7 data:

*r*(S) : -0.07 slope(S): -0.01

*r*(N) : +0.04 slope(N): +0.01

For wind speed vs. latitude, once again, both *r*-values and slopes are ~ 0. The only assessment from the plots would be that there is a greater proportion of cities in the southern hemisphere with higher wind speeds than in the northern hemisphere.

Graph 7(S, N)

A screenshot of a graph

Description automatically generated

Graph 8 data:

*r*(S) : +0.02 slope(S): +0.07

*r*(N) : -0.06 slope(N): -0.12

For plots of Cloudiness vs. latitude, the r-values and slopes are close to 0. There is an indication of a small negative slope in the northern hemisphere plot, which indicates that there are more cities with high cloudiness between 0° and 30° latitudes. This may or may not be of importance to vacationers.

Graph 8 (S, N)

A screenshot of a graph

Description automatically generated

Part II

In this section of analysis, all the city names along with the corresponding weather features, were obtained from Part I. From this data, a tourist/vacationer can select potential places of interest, by narrowing the range of temperature, humidity, cloudiness and wind speed data.

Using geodata from <https://myprojects.geoapify.com> the following map was plotted.

The map shows that there are more coastal cities than cities away from coasts. The size of each spot is determined by the humidity in the city.

A map of the world with different colored dots

Description automatically generated

By being selective about the weather variability, a much smaller number of cities suitable for a user as vacation spots were obtained.

This process reduces the number of cities of interest to a vacationer dramatically. The new number of cities is much less than the original number of over 500.

Moreover, not all the cities in the selection have hotel accommodations for tourist. If **only one** hotel were to be located within 10 km radius from a city center, the number of cities is reduced to less than 20.

The following map illustrates the reduced number of spots.

Hovering the mouse over each city, one can see the latitude, longitude, city name , country name, humidity and the name of a hotel.

A map of the world with different colored dots

Description automatically generated

**Conclusion**:

* From Part I, it can be observed that there are more cities in the northern hemisphere than in the southern hemisphere.
* A comparison between each weather feature between the northern and southern hemispheres shows the only significant correlation to be between **temperature** and latitude.
* However, from a tourist’s perspective, a lot of important information can be obtained from the initial graphs (1, 2, 3, 4) without the linear regression.
* Part II is a good exercise to find places suitable for vacation.

But by limiting the radius from the city center to 10 km, and by choosing only one hotel in the area, the selection may yield less satisfactory results than expected.