

FlexGen Internship Data & Analysis Task

SURBHI GARG

Task

Perform data evaluation and preparation for analysis, and conduct analysis of temperature for the presented dataset.

Objectives

- perform exploratory data analysis (your preferable tool, Excel or Jupyter)
- Build graphical representations of the analyzing dataset: weekly (min, max, avg).
- identify annual monthly seasonality and trends of the presented temperature observations based on the average temperature
- Identify the correlation between weekly average temperature and precipitation

About Data

There are two cities, San Juan and Iquitos, with weather/temperature data for each city spanning 5 and 3 years respectively. The data for each city have been concatenated along with a city column indicating the source: sj for San Juan and iq for Iquitos. Throughout the dataset, missing values have been filled as NaNs.

- **city** – Temperature observations locations
- **year** – Year of the temperature observation
- **weekofyear** – Week of year of the temperature observation
- **week_start_date** – Date of a week start of the temperature observation
- **station_max_temp_c** – Maximum temperature
- **station_min_temp_c** – Minimum temperature
- **station_avg_temp_c** – Average temperature
- **station_precip_mm** – Total precipitation
- **station_diur_temp_rng_c** – Diurnal temperature range

DATA ANALYSIS STEPS

Step 1: Importing python Libraries and Loading the Data

```
In [22]: # importing all necessary libraries
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns #for visualization
import matplotlib.pyplot as plt #for visualization
```

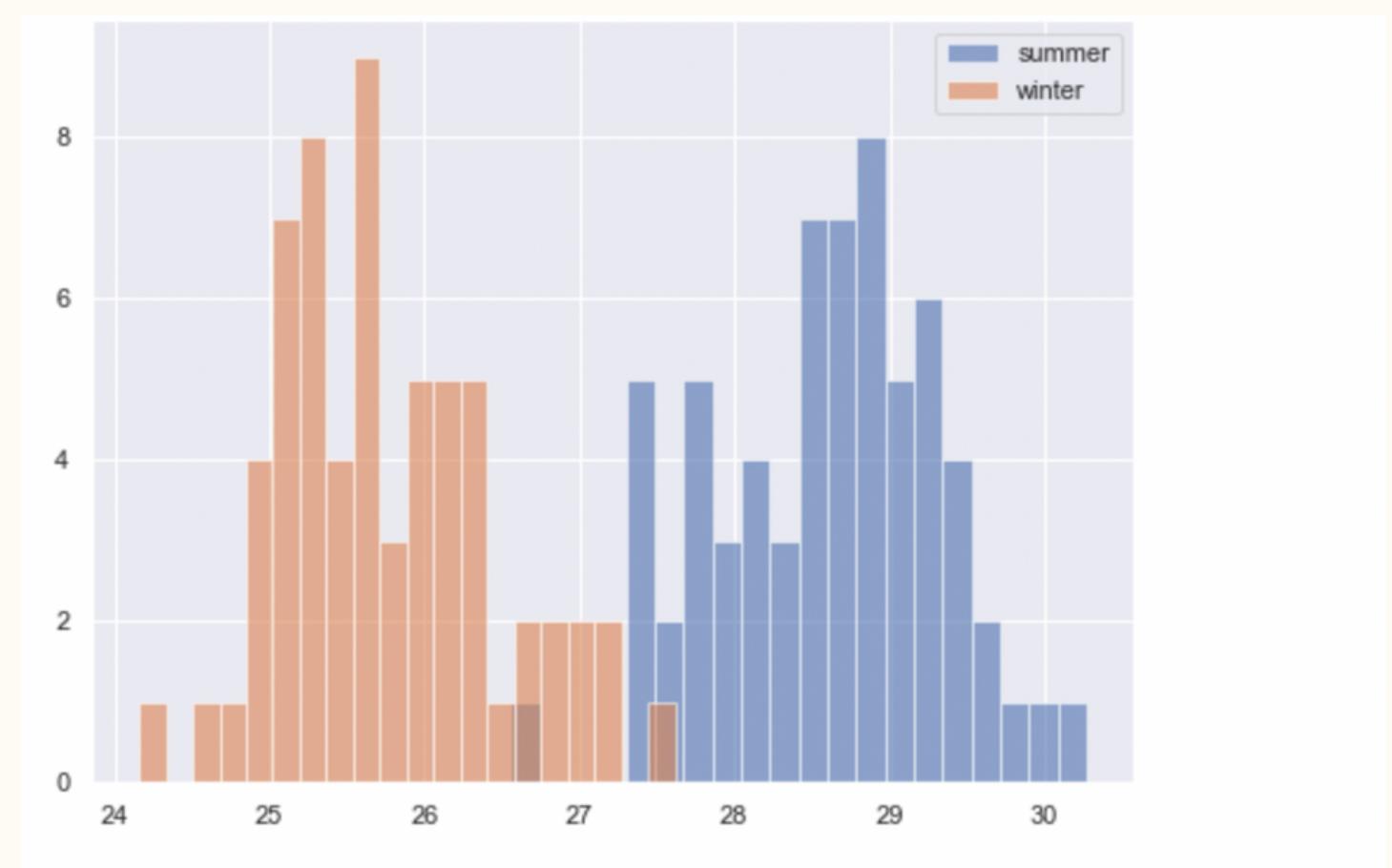
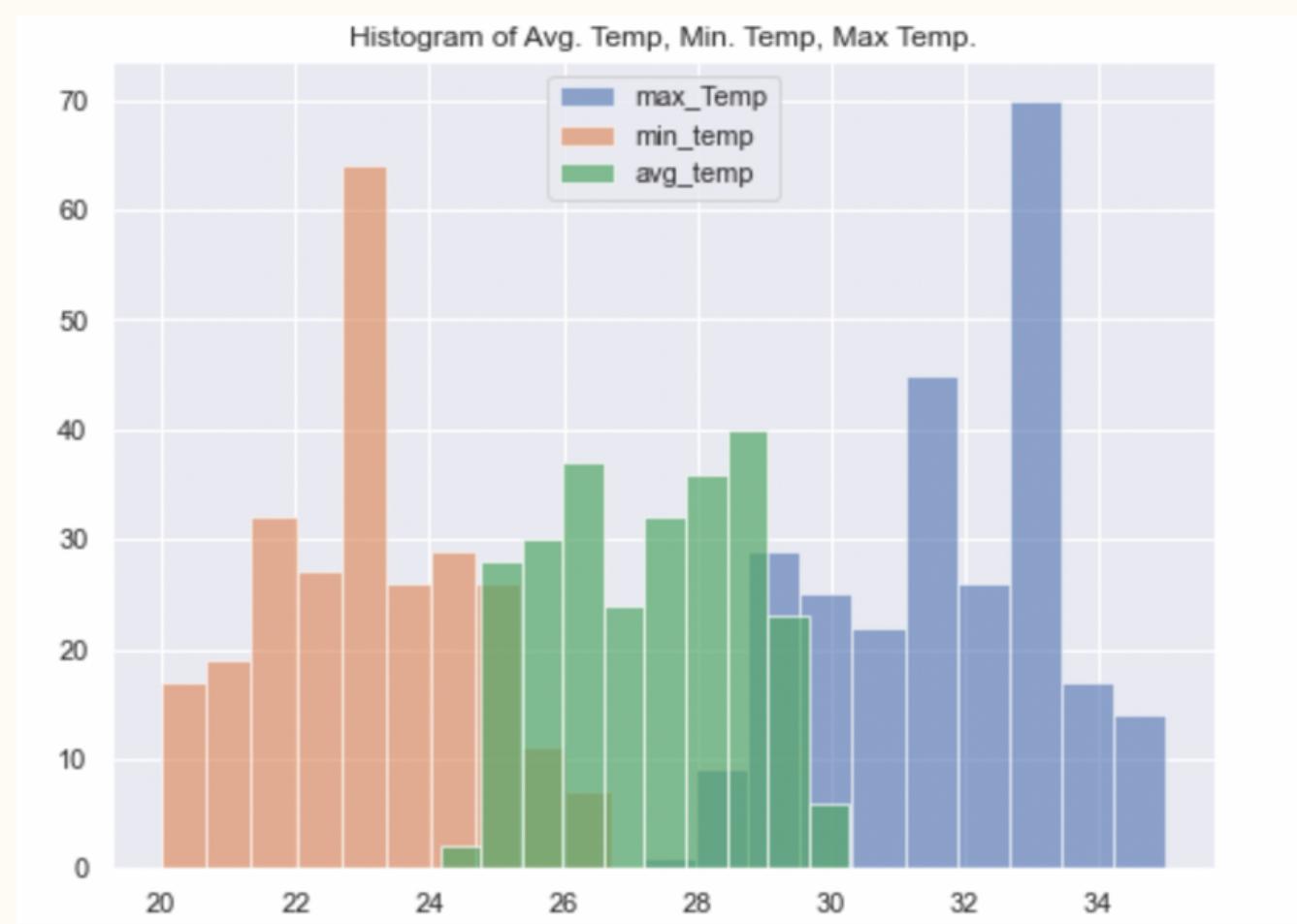
Importing Data

```
In [3]: # Importing data
df=pd.read_csv('flexgen_data_analysis_test.csv')
df.head()
```

Step 2: Data Preparation for Analysis

- Changing the data type of columns into the correct format – converted week_start_date to DateTime format
- Checked for duplicates – there were none
- replaced missing values using the interpolation method
- Separating the dataset set according to the cities of San Juan and Iquitos

EDA – San Juan



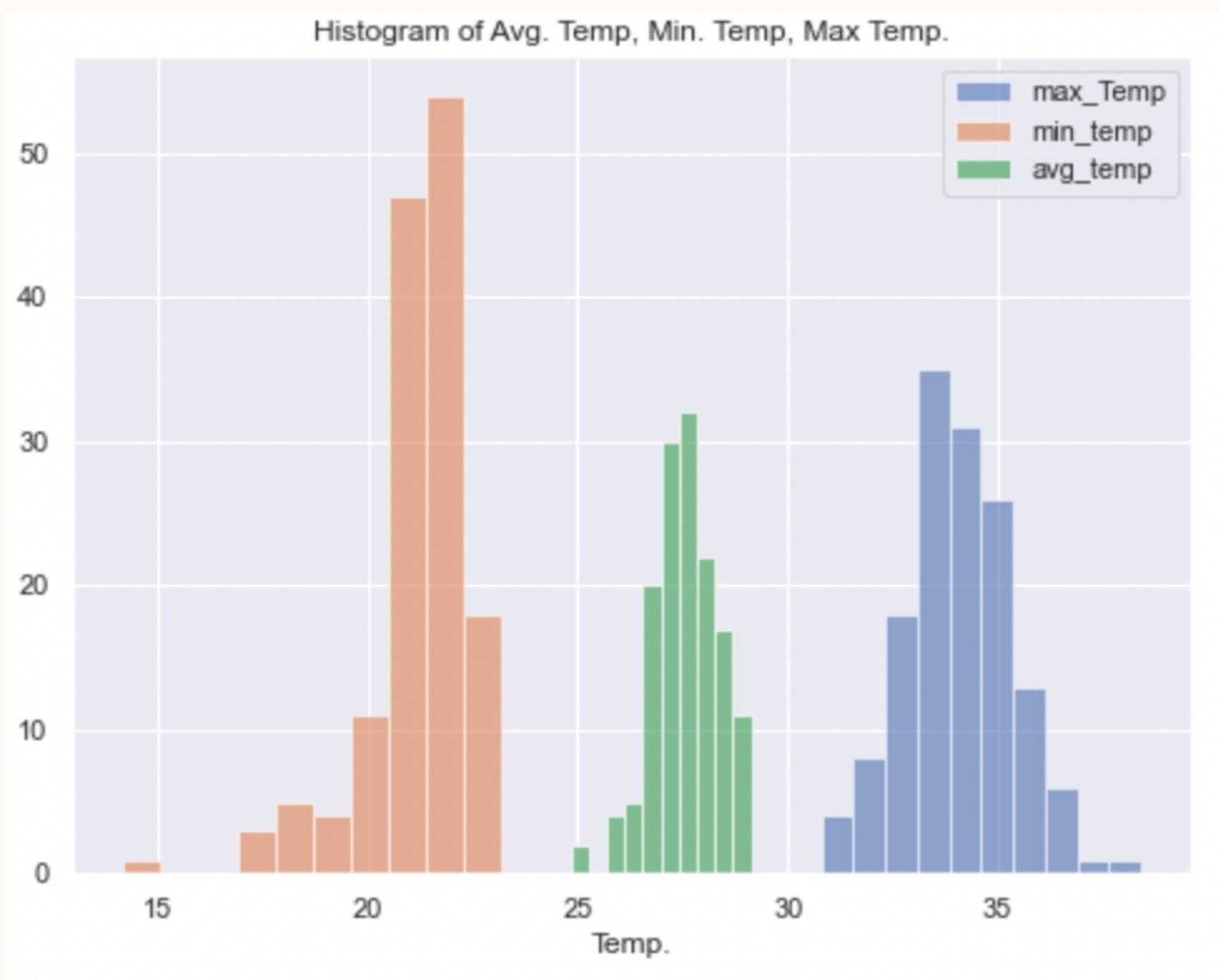
- According to Min temp. distribution, most temp. lies between 21 and 25-degree Celsius
- Avg temperature has a **bimodal behavior**, probably indicating peak temperature across the summer and winter seasons
- According to Max temp. distribution, most temp. points lies between 30 and 33 degrees Celcius
- Divided data into summer and winter months to study bimodal nature of Avg Temp graph (summer ~ June- Aug ~ **6 to 8 Months** & Winter ~ December to Feb ~ **12 to 2 months**)
- For Winter months, most temp. points lie between 25 to 27 degrees celcius with a peak at 26
- For Summer months, most temp points lie 28 to 30 degree Celsius with the peak at 29

EDA – San Juan

	station_max_temp_c	station_min_temp_c	station_avg_temp_c			
	min	max	min	max	min	max
months						
Jan	27.2	31.7	20.6	22.8	24.157143	26.942857
Feb	28.3	32.8	20.0	22.2	24.828571	26.757143
Mar	28.3	33.9	20.0	22.8	24.785714	28.042857
Apr	28.9	33.9	20.6	23.9	25.057143	27.914286
May	29.4	34.4	20.6	24.4	25.942857	28.528571
Jun	30.6	35.0	21.1	26.1	26.557143	30.000000
Jul	31.1	35.0	22.8	26.1	27.328571	30.271429
Aug	31.1	34.4	22.8	26.7	27.885714	29.471429
Sep	31.7	35.0	22.2	26.1	26.828571	29.728571
Oct	31.1	33.9	22.8	26.1	27.014286	29.728571
Nov	29.4	34.4	22.2	25.0	26.157143	29.400000
Dec	28.3	32.2	20.0	23.9	24.657143	27.614286

- Studying the Maximum and Minimum Temp. across months
- The least max. temp was 27.2 in Jan and the highest max temp was 35 in June, July, and Sept.
- least min temp was 20 in Dec, Feb, Mar and the highest min temp was 22.8 in July, Aug, and Oct
- least avg temp was 24.1 Jan and the highest max temp was 30.27 in July
- The temp. ranges indicating the different seasons in the city with winter being around Dec, Jan, and Feb and Summer being around June, Jul, Aug, Sept.

EDA – Iquitos



- Min temp graph is left-skewed with most of the min temp > 20 degrees Celsius
- Avg temp graph is left-skewed with most of the temp b/w 26 and 29 degree Celsius
- Max temp is somewhat normally distributed with most of the temp around 34 degrees Celcius

EDA - Iquitos

months	station_max_temp_c		station_min_temp_c		station_avg_temp_c	
	min	max	min	max	min	max
Jan	31.2	36.0	20.8	23.2	26.450	28.500000
Feb	32.4	34.9	20.5	22.4	26.850	28.350000
Mar	32.2	35.2	20.4	22.8	26.600	28.366667
Apr	32.8	35.0	18.2	23.1	26.100	28.700000
May	30.8	34.6	19.4	22.5	26.300	28.450000
Jun	32.2	34.4	17.8	22.2	24.840	27.520000
Jul	31.4	34.0	14.2	21.6	25.100	27.440000
Aug	33.4	35.2	17.4	21.2	26.275	28.750000
Sep	33.8	36.2	17.8	21.6	27.100	28.866667
Oct	33.3	38.4	19.8	22.8	26.800	29.133333
Nov	32.2	37.4	20.0	22.5	26.650	28.833333
Dec	33.3	36.2	20.5	22.5	27.080	28.766667

- least max temp was 30.8 in May and the highest max temp was 38.4 in Oct.
- least min temp was 14.2 in Jul and the highest min temp was 23.2 in Jan
- least avg temp was 24.8 Jun and the highest max temp was 29.1 Oct
- The weather in Iquitos is opposite of San Juan as the maximum and minimum temperature occurs in different months

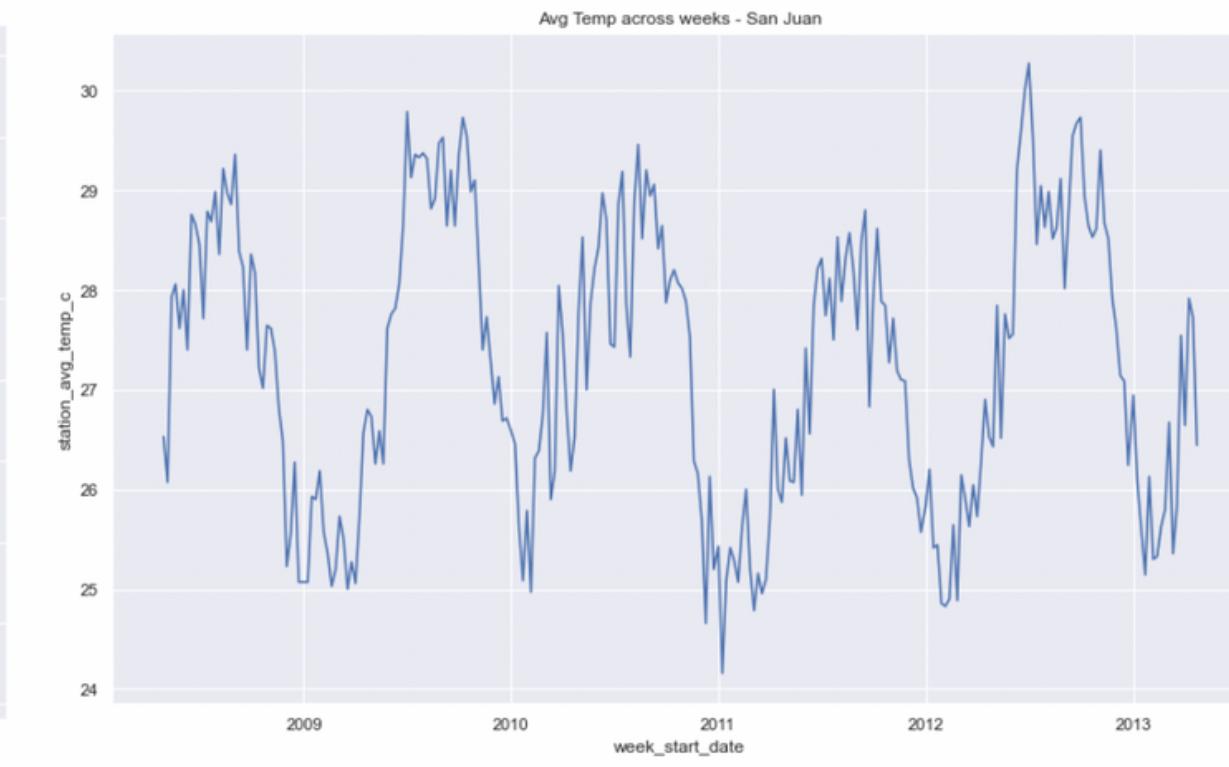
Graphical representation of weekly temp.- San Juan



Max Temp. across weeks



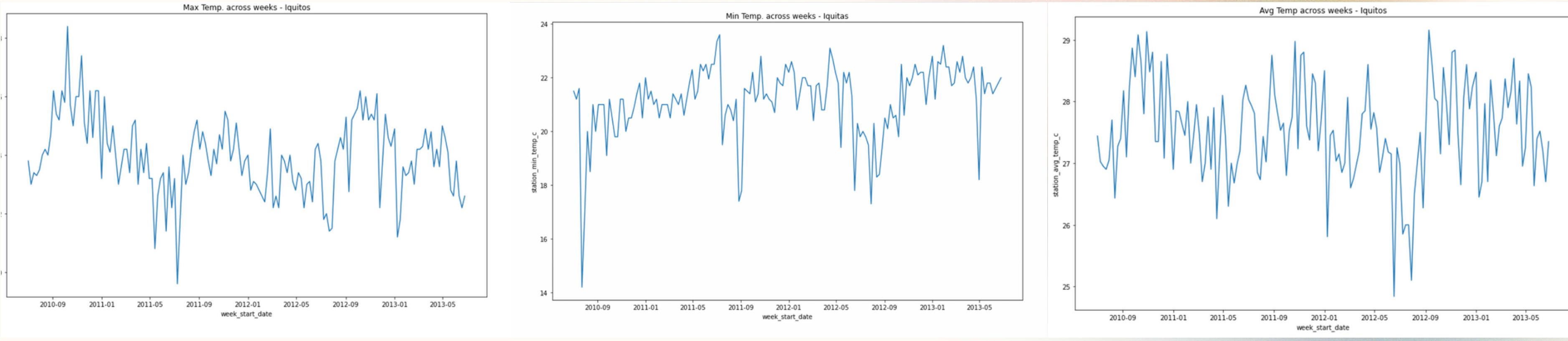
Min Temp. across weeks



Avg Temp. across weeks

- Across all three graphs, we can see there is a horizontal trend along with a seasonality with a period of 1 year
- At the start of every year, there is a drop in the temperature which increases and achieves a peak around mid-year indicating the summer and winter seasons

Graphical representation of weekly temp.- Iquitos



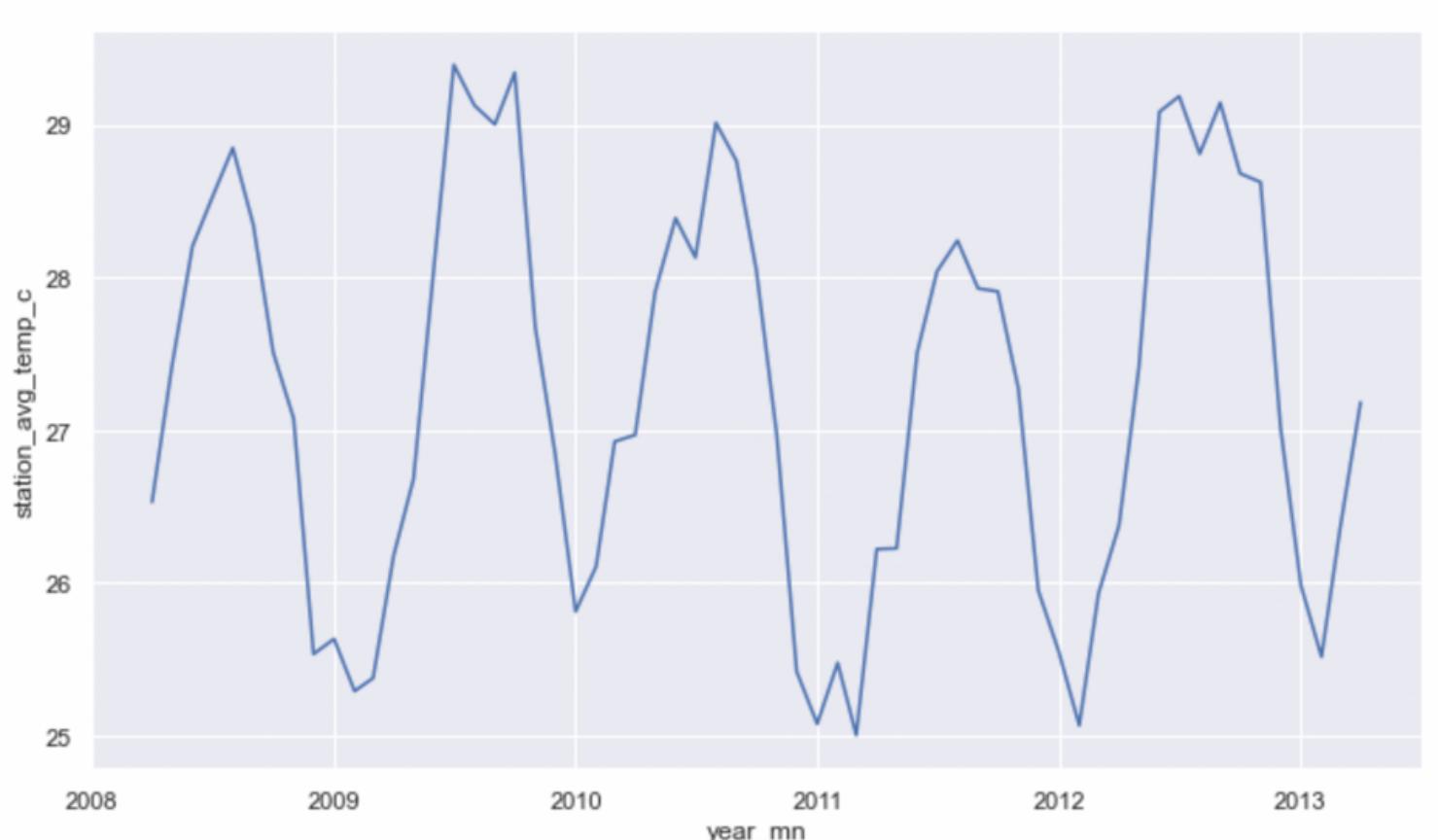
Max Temp. across weeks

Min Temp. across weeks

Avg Temp. across weeks

- A downward trend can be observed in Max and Avg. temp graph
- An upward trend can be observed in Min Temp. graph.
- At the start of every year, there is an increase in the temperature which decreases and achieves the lowest temperature around mid-year indicating the summer and winter seasons (opposite to San Juan)

Annual Monthly Avg Temp Analysis- San Juan

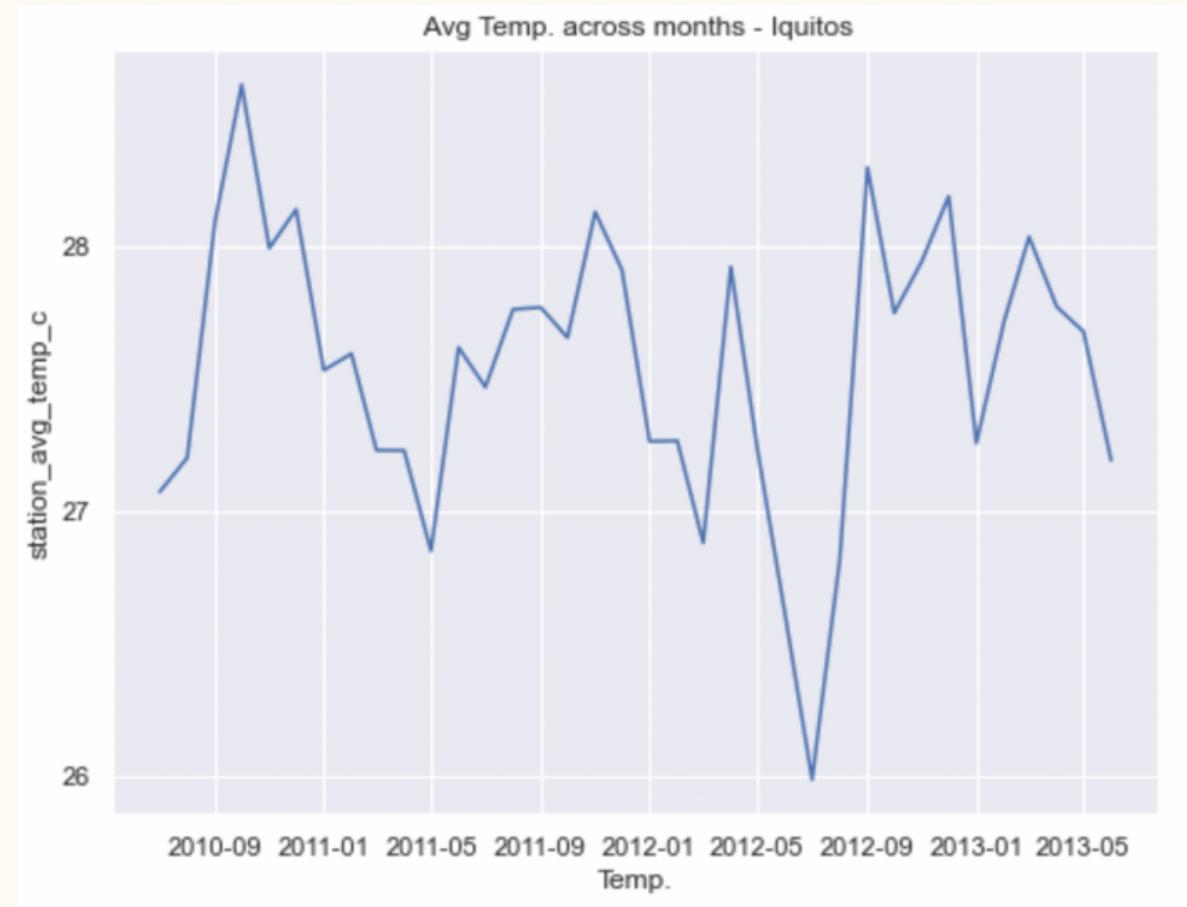


Avg Temp. across Months

year	month	station_precip_mm			
		min	max	mean	
33	2011	1	2.5	24.1	14.360
34	2011	2	6.1	25.1	15.225
35	2011	3	1.0	11.6	4.500
36	2011	4	0.3	23.1	9.380
37	2011	5	0.0	89.5	53.550
38	2011	6	61.3	200.6	110.275
39	2011	7	6.4	99.6	42.440
40	2011	8	20.3	205.8	97.575
41	2011	9	21.5	153.1	59.225
42	2011	10	10.4	58.0	37.820
43	2011	11	15.2	39.0	27.025
44	2011	12	27.0	93.7	60.450

- there is a horizontal trend along with a seasonality with a period of 1 year
- Avg. temp around the Summer of 2011 was the lowest compared to other years
- Avg. Rain during the mid months of 2011 was the highest, probably the reason behind the lower temperature

Annual Monthly Avg Temp Analysis- Iquitos

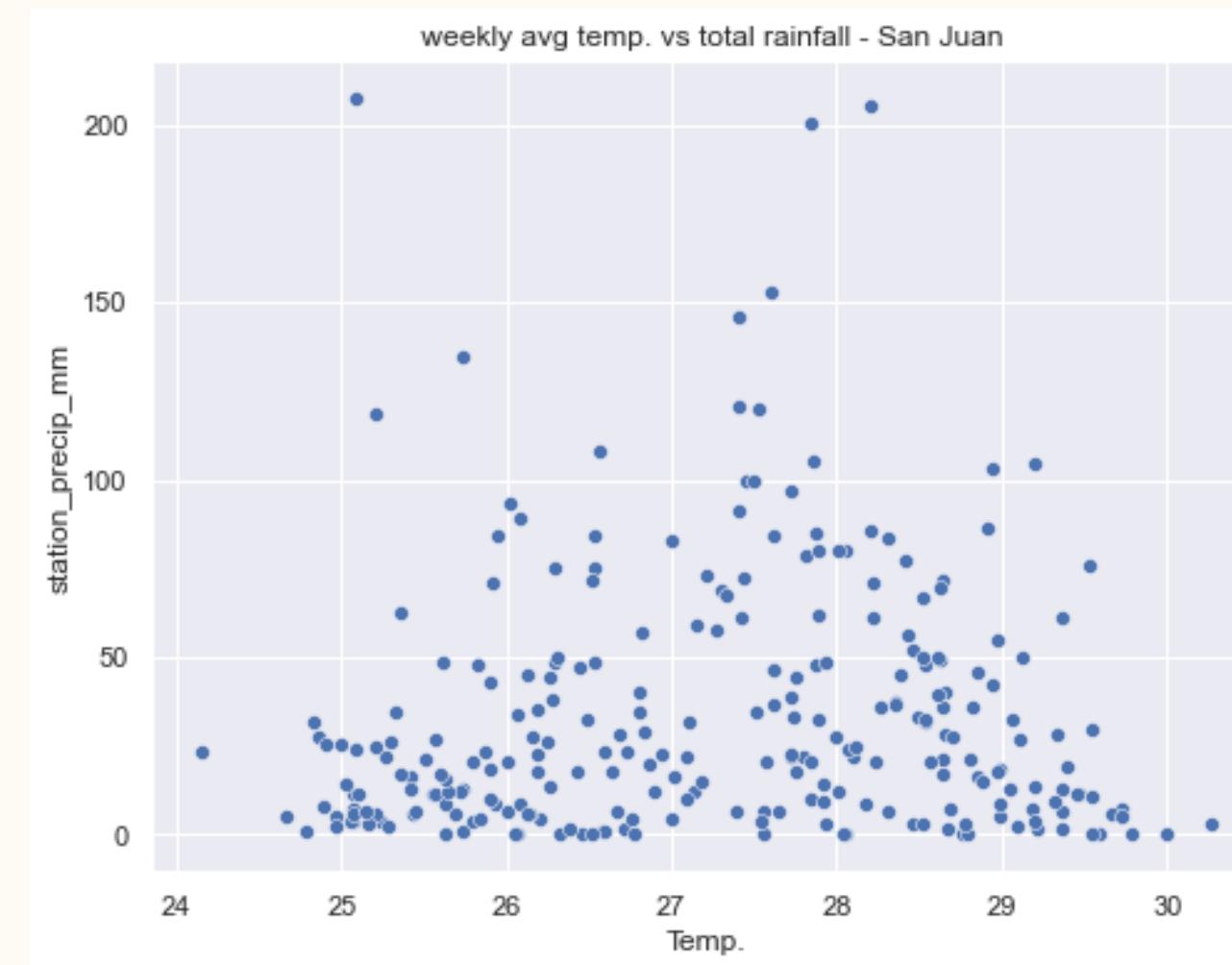


Avg Temp. across Months

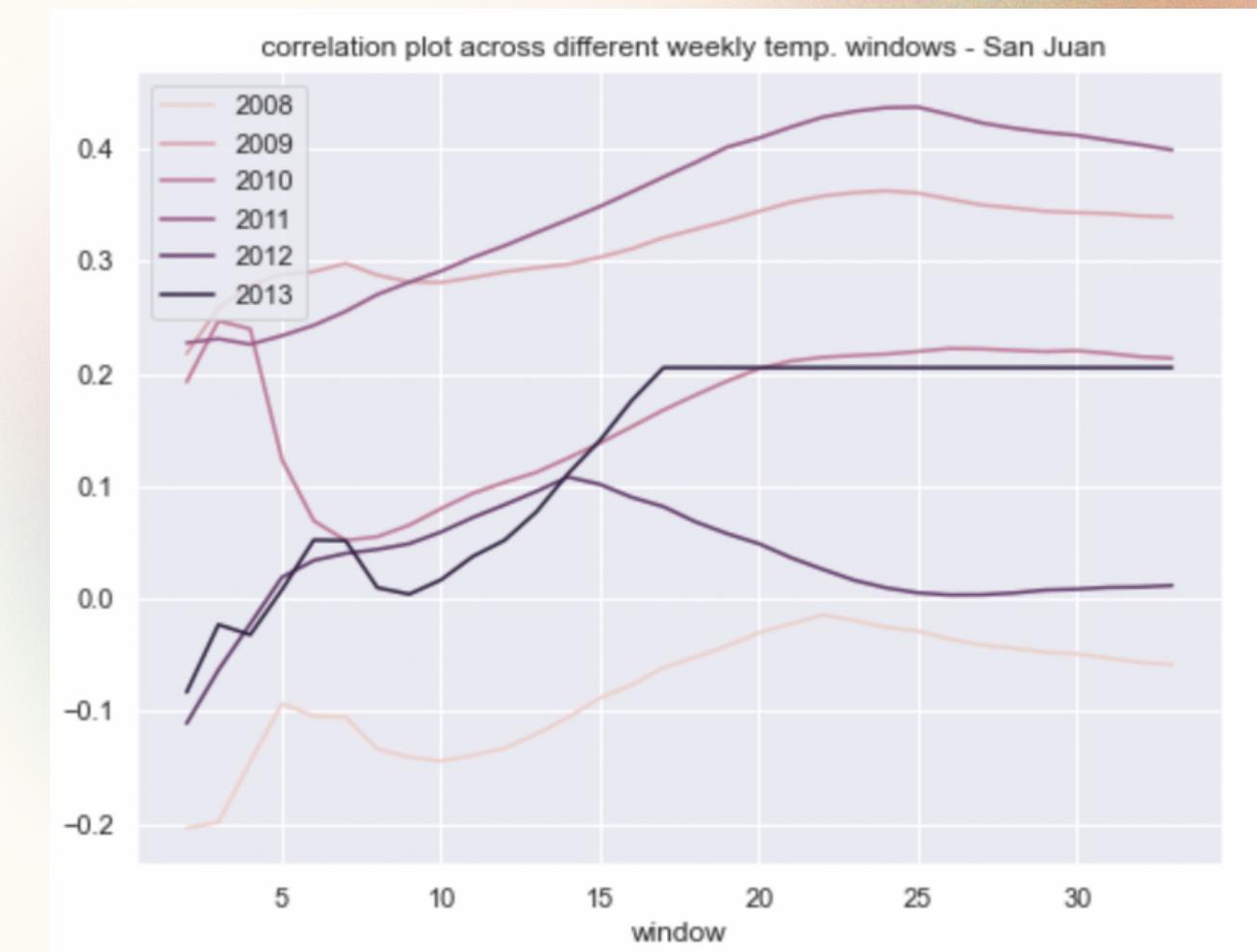
year	month	station_precip_mm		
		min	max	mean
18	2012	1	14.7	132.30
19	2012	2	23.0	94.50
20	2012	3	49.2	138.10
21	2012	4	7.9	212.00
22	2012	5	9.1	67.90
23	2012	6	3.8	68.20
24	2012	7	9.1	37.00
25	2012	8	0.0	59.85
26	2012	9	6.1	28.90
27	2012	10	14.0	51.50
28	2012	11	7.1	64.00
29	2012	12	3.0	50.00
				26.0750

- there is a Downward trend along with a seasonality with a period of 1 year
- Avg. temp around the Summer of 2012 dropped the most compared to other years
- The drop in temp. may be due to the advent of winter season

Correlation Analysis- San Juan



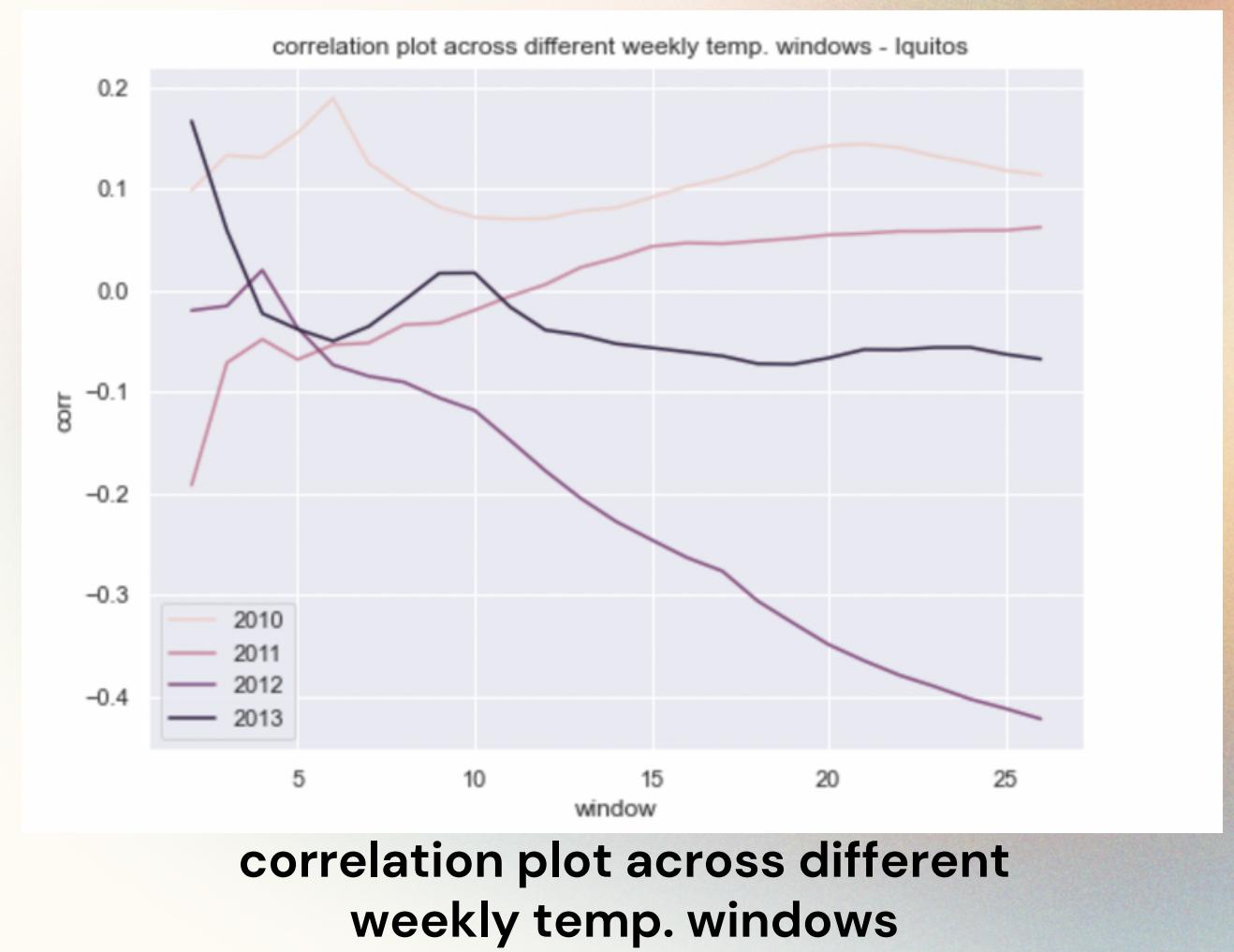
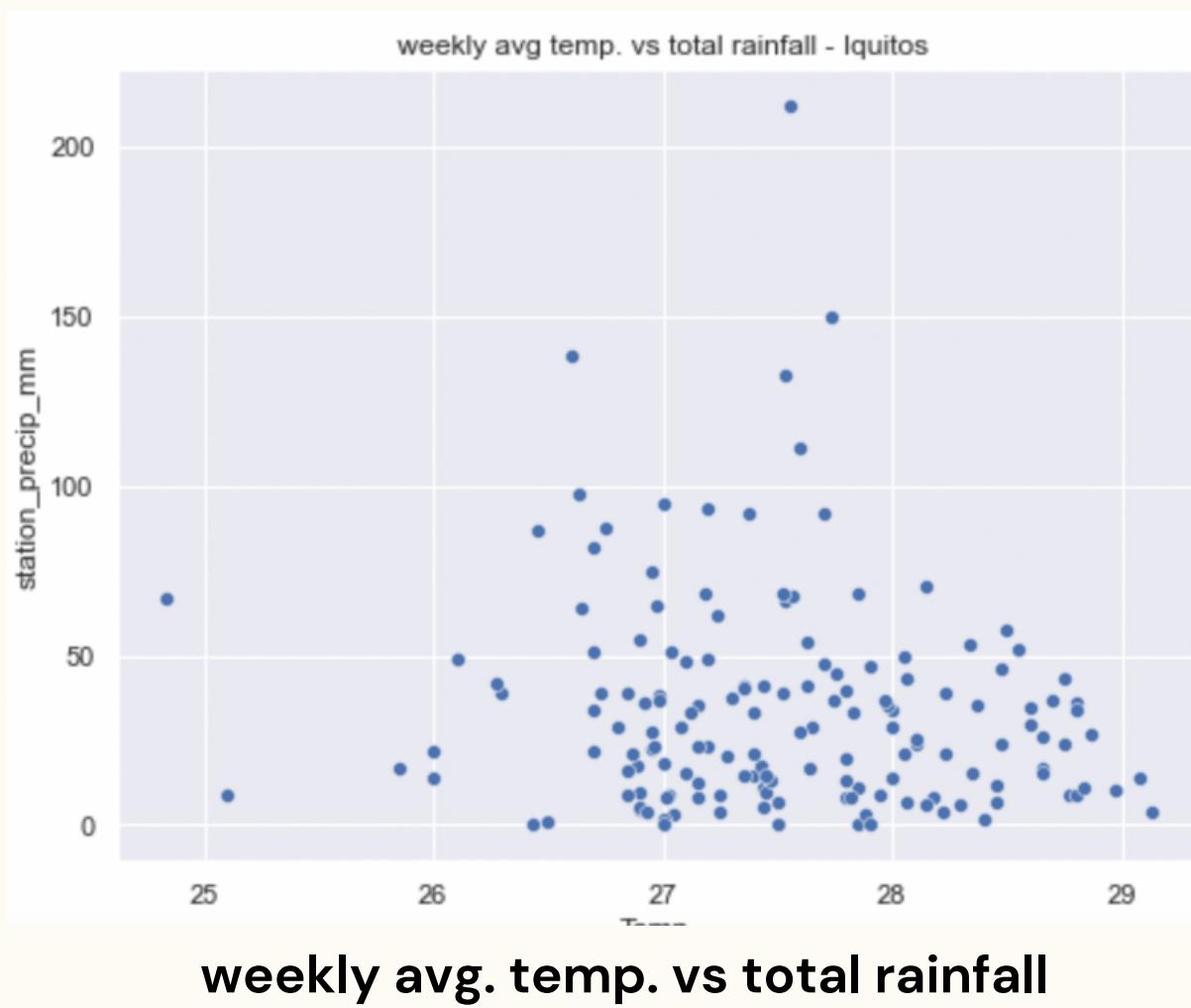
weekly avg. temp. vs total rainfall



correlation plot across different weekly temp. windows

- Corr value weekly avg. temp. with total rainfall: 6.4 %
- correlation value increases as window size increase until a threshold

Correlation Analysis- Iquitos



- Corr value weekly avg. temp. with total rainfall: -12.8 %
- correlation value pattern differs. the corr. value decreases as window size increase until a threshold

Key Takeaways

- Temp. Fluctuations in both cities indicate they belong to different hemispheres
- According to Temp. variation chart, we can indicate San Juan most probably belongs to the Northern hemisphere and Iquitos to the southern hemisphere as the months with min. and max. temperature are opposite in both the cities
- For San Juan, Min., Max, Avg. Temp. follows a seasonal horizontal Trend. For Iquitos Downwards and upwards trends are observed across the three temp. values
- For San Juan, Monthly Avg. Temp. follows a seasonal horizontal Trend. For Iquitos seasonal and Downward trend is observed
- For San Juan, week avg. rainfall isn't correlated to the weekly rainfall but rather to the avg. temp. of the previous week.

For Iquitos, the weekly avg. rainfall is negatively weekly correlated to the weekly rainfall. the corr. value decreases as window size increase until a threshold

THANK YOU !!