

A Statistical Analysis of the Accuracy of 5-day Forecasts in Predicting Hot vs Cold Temperatures

The Code

We had to use old python web scraping packages as the newer ones were not capable of running on such an old system

```
import sys
import urllib.request
import json
from json import loads
import csv
import datetime
import os
```

```
city_name =
str(sys.argv[1])
```

```
if city_name ==
"Winnipeg":#12pm
location_code =
str(48989)
else:
location_code =
str(113487)

url_forecast="http://dat
aservice.accuweather.com
/forecasts/v1/daily/5day
/"+location_code+".json?
APIkey"
```

```
now=datetime.datetime.no
w()
date_now=now.strftime("%
d/%m/%Y")
date_file=now.strftime("
%d%m%Y")
time_now=now.strftime("%
H:%M:%S")
time_file=now.strftime("
%H%M%S")
print(str(date_now)+" "+s
tr(time_now))
```

```
req=urllib.request.Reque
st(url_forecast)
```

```
With urllib.request.url
open(req) as response:
page= response.read()
```

```
cont_forecast=loads(page
.decode('utf8'))
```

```
date_forecast = []
min_temp = []
max_temp = []
```

```
for k in range(5):
```

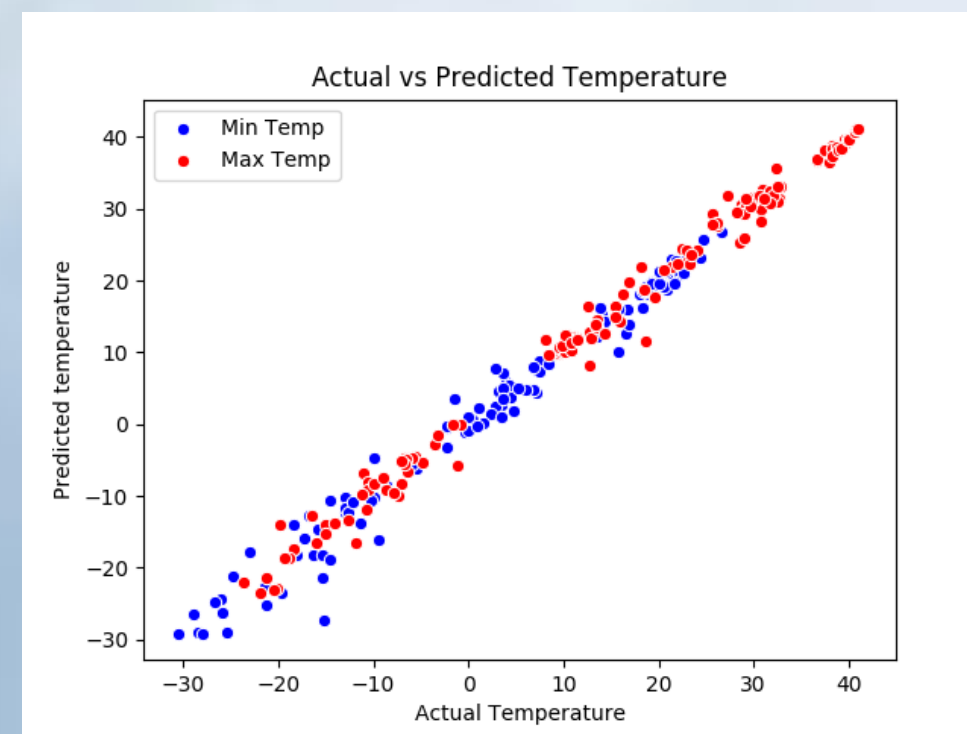
```
date_forecast.append(con
t_forecast['DailyForecas
ts'][k]['Date'])
```

```
min_temp.append(cont_for
ecast['DailyForecasts'][
k]['Temperature']['Minim
um']['Value'])
```

```
max_temp.append(cont_for
ecast['DailyForecasts'][
k]['Temperature']['Maxim
um']['Value'])
```

```
orig_data_file_name='/ho
me/python_code/weather_d
ata_'+str(city_name)+'.c
sv'
```

Statistical Analysis



This graph is a representation of our total data collected from both warm and cold climates. It shows that the coldest temperatures are harder to predict and have a larger difference between the prediction and the real temperatures. As the temperatures rise the predictions become more accurate and the warmest temperatures recorded show very little deviation from the predicted temperatures.

Introduction

This project is a statistical analysis of the 5-day temperature forecast for 7 cities (Winnipeg, Kinshasa, Rome, Astana, Sarh, Buenos Aires and Houston), over 20 days in January 2019 versus the actual recorded temperature for those 20 days.

Our group decided to choose this topic as we were both interested in it. Our interest grew while we were both on holidays last summer, one of the group members was visiting a cold climate and the other a warm climate. We noticed that the temperature predictions in the warm climate tended to be much more accurate than the temperature predictions for the cold climate. What also intrigued us was that when we researched it online there was no theory about temperature predictions being more accurate in warm climates so we wanted to figure it out for ourselves.

Collecting The Data

We decided that the most accurate weather data site would be the optimum place to get our data. After doing some research we found that AccuWeather is known to be the most accurate in predicting weather forecasts.

With the cities we had chosen being scattered around the globe, we needed to take time zones into consideration. For our data to be most accurate we needed to collect it at 6am in each of the cities. This would have made it very difficult for us to record the data manually because we would have had to check AccuWeather during the night too. Collecting it by hand would also have inevitably led to many small errors within the data itself.

Therefore we decided that it would be best to have a computer do this work for us. Using Python, we designed a web scraping programme to collect data from the AccuWeather site daily, and at specific time. This data was then directly exported to Excel where we could easily edit it. This programme ran all day and night on a NSLU2.



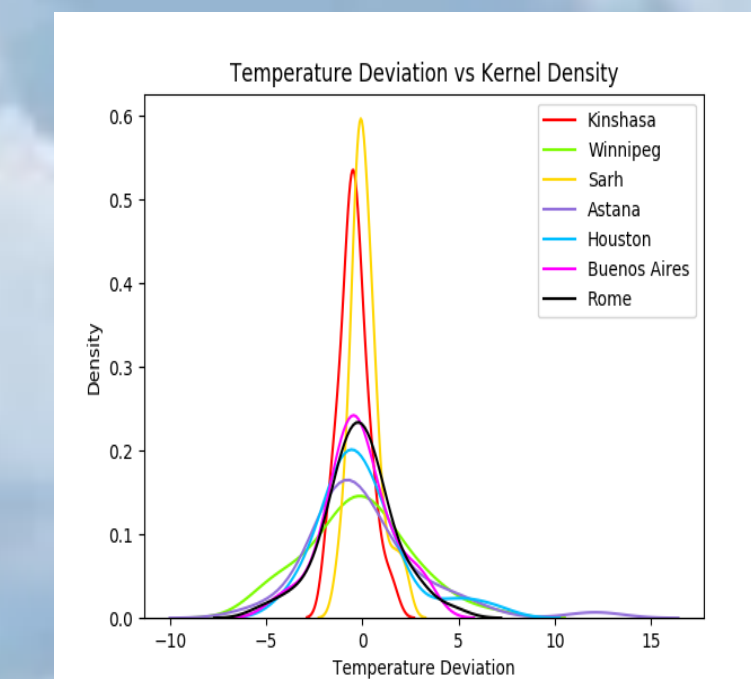
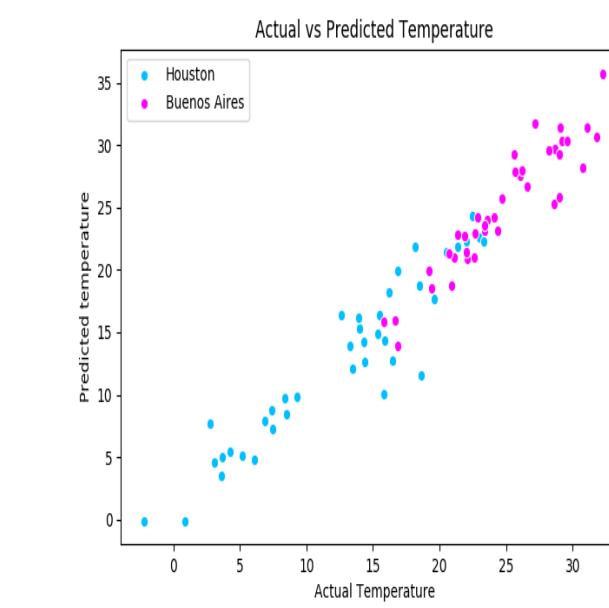
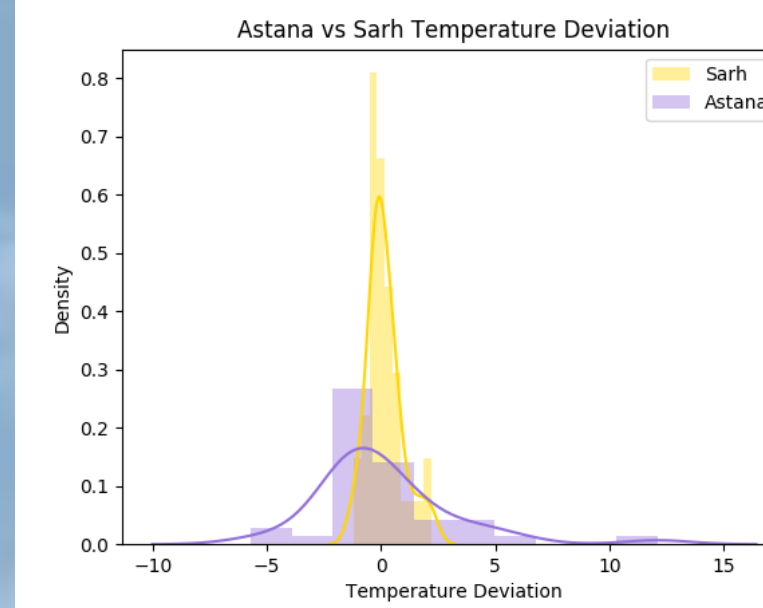
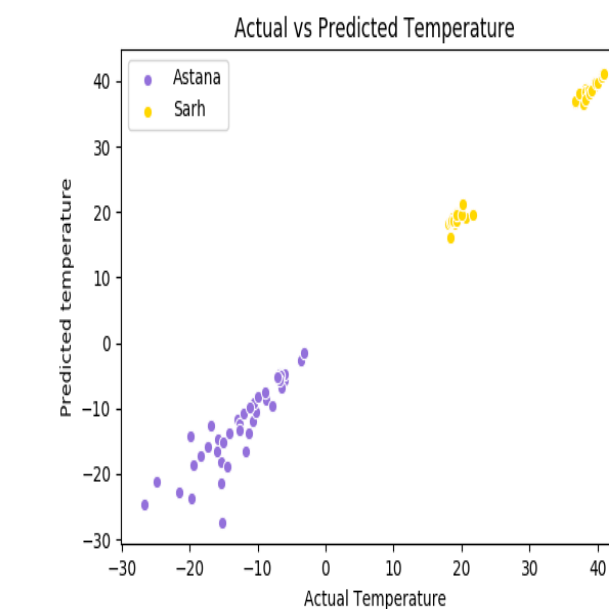
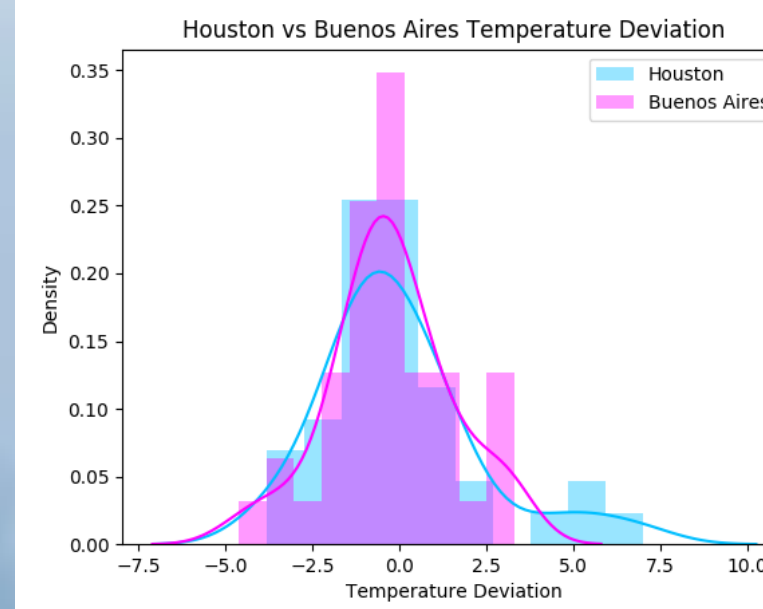
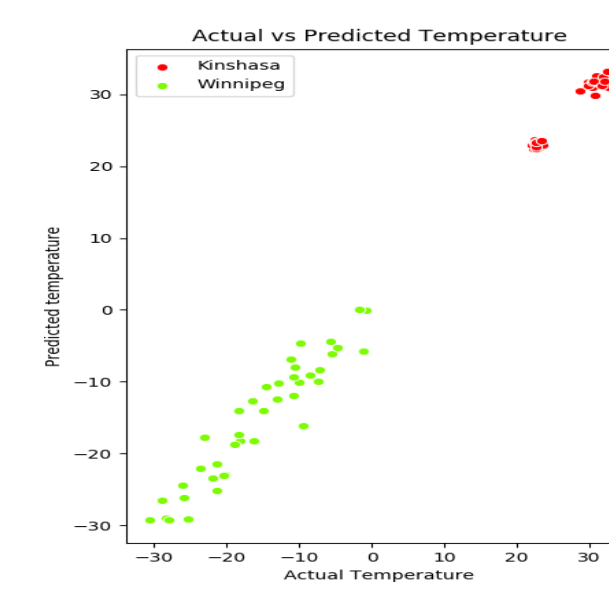
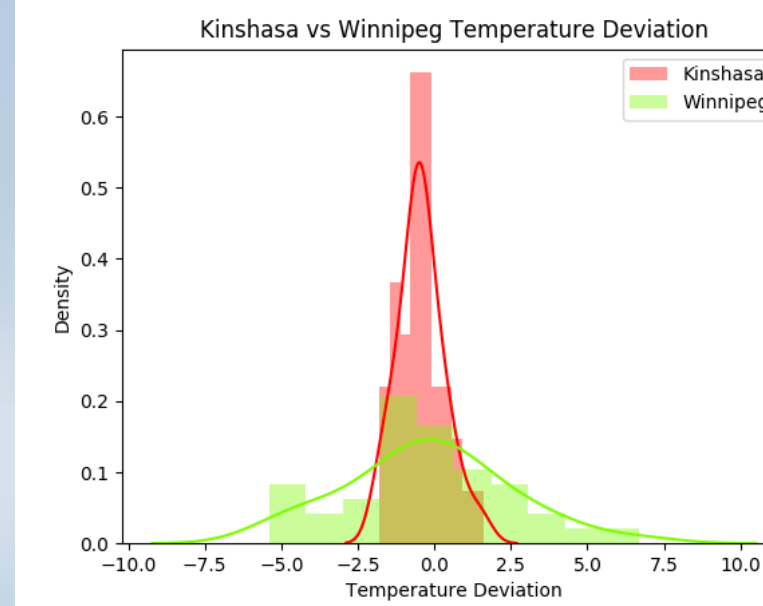
This table shows the max temperature data for the 20 days

The Data

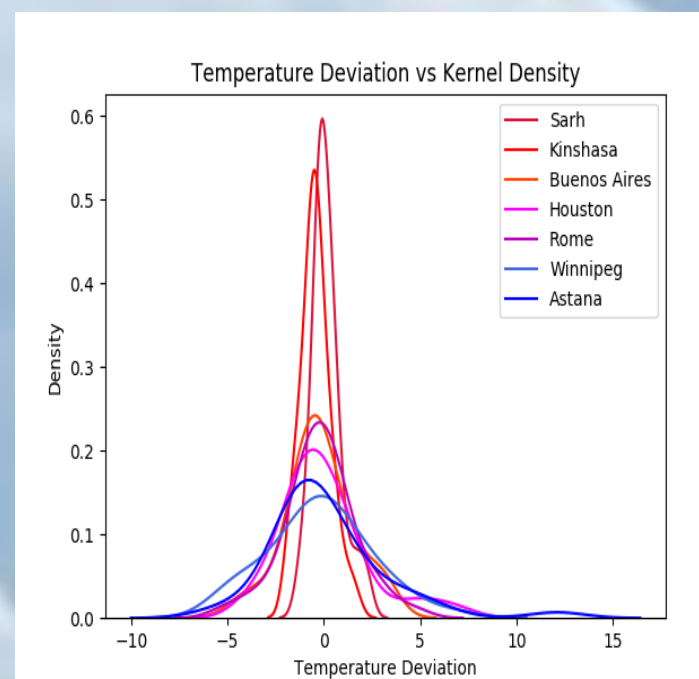
Kinshasa		Rome		Houston		Winnipeg		Astana		Buenos Aires		Sarh	
Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual
32.7	31.5	11.1	12.1	15.9	14.4	-0.8	-0.1	-19.4	-18.6	25.6	29.3	38	37.9
30.4	31.4	10.4	12.5	18.2	21.9	-1.1	-1.1	-19.8	-14.1	28.7	29.7	38.2	38.8
28.7	30.5	9.5	9.5	21.4	21.9	-5.7	-4.4	-14.1	-13.7	30.8	28.2	38.2	38.4
30.6	31.8	9.7	10.1	23	22.6	-1.7	0	-10.6	-9.2	26.1	27.6	37.4	37.6
30.3	31	9.8	10.3	22.5	24.4	-7.4	-9.9	-11.2	-9.7	31.8	30.7	36.7	36.8
30.7	29.9	8.1	11.8	16.2	18.2	-15	-14.1	-9.9	-8.3	29.2	30.4	36.7	36.9
29.9	31.7	13.6	14.6	15.5	16.4	-8.6	-9.1	-16	-16.5	24.1	24.2	37.4	38.2
30.9	32.6	12.7	12.9	16.9	19.9	-11.1	-6.9	-15	-15.2	26.2	28	38.3	38.4
29.9	31.2	12.9	12	13.3	13.9	-4.8	-5.3	-12.7	-13.4	28.6	25.3	39.5	39.1
31.3	31.7	13.5	13.4	14.4	12.6	-7.1	-8.3	-7.8	-9.6	25.7	27.9	39.6	39.7
31.1	31.6	11.5	12	13.3	13.9	-18.4	-17.4	-6	-4.8	31.1	31.4	38.7	38.6
30.3	31.7	9.6	10.8	23.3	22.3	-21.9	-23.5	-6.7	-4.7	22.9	24.2	38	36.4
32.8	33.1	12.7	8.1	22	22	-23.6	-22.1	-6.4	-6.7	23.4	23.6	38.3	37.3
30.6	31.8	10.2	10.1	18.6	11.6	-10.8	-11.9	-6.6	-5	29	29.3	38.9	38.1
32.6	31	10.5	11.8	19.6	17.7	-16.5	-12.7	-3.6	-2.7	28.2	29.6	39.2	38.5
31.8	32.5	10.8	10.4	20.5	21.5	-10.6	-8	-3.2	-1.5	29.6	30.4	39.8	39.7
31.7	31.2	9.9	10.9	8.4	9.7	-21.3	-21.5	-8.9	-7.5	29	25.9	40.1	39.7
32.1	31.8	10.1	12.4	15.4	14.9	-20.2	-22.9	-6.8	-5.6	27.2	31.8	40.6	40.7
32.9	33.1	10.8	11.3	12.6	16.4	-20.4	-23.1	-7	-5.2	29.1	31.4	40.8	41.2
32.5	33.2	11.5	11.7	18.5	18.8	-18.9	-18.7	-11.8	-16.5	32.3	35.7	41	41.2

Temperature Deviation

The deviation was found by taking the predicted temperature and subtracting the actual temperature from it. They encompass the temperature deviation for both the min and max temperature predictions.



These graphs show the same data, with the left one showing hotter cities in red, milder ones in pink, and cold ones in blue



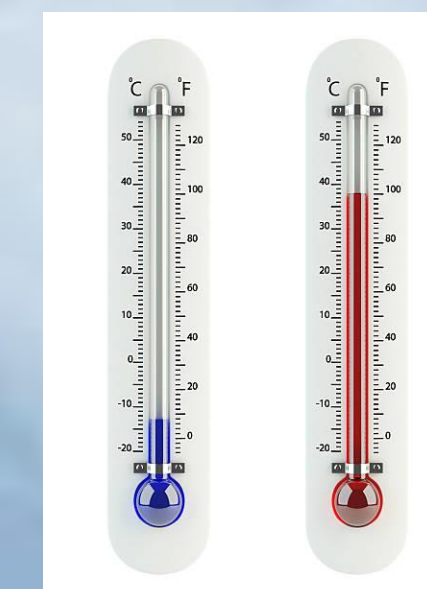
The Cities

The cities we decided to use are Kinshasa (Dem. Rep. of Congo), Winnipeg (Canada), Sarh (Chad), Astana (Kazakhstan), Houston (USA), Buenos Aires (Argentina) and Rome (Italy). Of these seven, six of them are paired into groups (Winnipeg and Kinshasa, Sarh and Astana, Houston and Buenos Aires). These pairs were designed to make the cities more comparable. Because there are many factors that go into making a forecast accurate, we tried to find cities that had similar traits (altitude, distance from a sea/ocean, mountains close by). The pairs consist of one city that's generally cold this time of year and one that's generally warm. We believed that with only these pairs we would've been left with gaps between the extreme cold/hot, which is why we decided to include Rome, which at this time of year has a mild climate.



	Avg Jan Min Temp (°C)	Avg Jan Max Temp (°C)
Kinshasa	22	30
Winnipeg	-18	-10
Astana	-21	-11
Sarh	17	36
Buenos Aires	21	29
Houston	7	16
Rome	3	12

We used cities from both the northern and southern hemispheres to get more variety in our hot and cold climates



Conclusion

From this analysis we can conclude that:

- Hotter temperatures are easier to predict than colder ones
- However, the statement is true mostly for severe hot/cold temperatures, with there being a visible middleground around -5 to 20 where the prediction accuracy seems to be very similar.