

Working with Document Stores

```
In [ ]: # imports
import pymongo
import pandas as pd
from plotly.subplots import make_subplots
import plotly.express as px
import plotly.graph_objects as go
import seaborn as sns
import matplotlib.pyplot as plt
```

Connect to Mongo

```
In [ ]: # connect to mongo
client = pymongo.MongoClient("mongodb://127.0.0.1:27017/")

# connect to mars database
db = client["mars"]

# connect to weather collection
collection = db["weather"]
```

Tutorial

Query is first line of code in cell

Output shown as head of dataframe or printed statement

Q1. What are the ten most recent observations?

```
In [ ]: # query the collection for data sorted by terrestrial_date descending and limit
data = collection.find().sort("terrestrial_date", -1).limit(10)

# create a dataframe from the data
df = pd.DataFrame(data)

# drop the _id column
df.drop(columns=["_id"], inplace=True)

# show the dataframe
df.head(100)
```

Out []:

	id	terrestrial_date	sol	ls	season	min_temp	max_temp	pressure	wind_speed	atm
0	3476	2022-11-25	3663	344	Month 12	-72	-44	817	None	
1	3474	2022-11-24	3662	343	Month 12	-72	-9	800	None	
2	3478	2022-11-23	3661	342	Month 12	-73	-8	797	None	
3	3477	2022-11-22	3660	342	Month 12	-72	-10	798	None	
4	3479	2022-11-21	3659	341	Month 12	-72	-12	798	None	
5	3475	2022-11-20	3658	341	Month 12	-71	-16	798	None	
6	3472	2022-11-19	3657	340	Month 12	-71	-6	797	None	
7	3471	2022-11-18	3656	340	Month 12	-70	-8	800	None	
8	3473	2022-11-17	3655	339	Month 12	-70	-7	797	None	
9	3470	2022-11-16	3654	339	Month 12	-70	-8	796	None	

Q2. What are the most recent observations where the difference in temperature between the min and max temperature is greater than 80 degrees?

```
In [ ]: # query the collection for where the difference between the max and min temperature is greater than 80 degrees
data = collection.find({"$expr": {"$gt": [{"$subtract": ["$max_temp", "$min_temp"]}]} })

# create a dataframe from the data
df = pd.DataFrame(data)

# drop the _id column
df.drop(columns=["_id"], inplace=True)

# show df
df.head(100)
```

Out[]:

	id	terrestrial_date	sol	ls	season	min_temp	max_temp	pressure	wind_speed	a
0	2851	2021-01-04	2991	342	Month 12	-100	-4	813	None	
1	2704	2020-08-06	2844	252	Month 9	-99	-1	889	None	
2	2665	2020-06-27	2805	227	Month 8	-99	1	859	None	
3	2666	2020-06-25	2803	225	Month 8	-100	1	854	None	
4	1991	2018-06-06	2073	187	Month 7	-74	7	756	None	
5	1361	2016-08-25	1441	210	Month 8	-74	10	828	None	
6	1333	2016-07-26	1412	192	Month 7	-78	4	784	None	
7	706	2014-09-29	763	204	Month 7	-78	6	817	None	
8	703	2014-09-28	762	203	Month 7	-75	7	814	None	
9	704	2014-09-25	760	202	Month 7	-70	11	810	None	
10	702	2014-09-24	759	201	Month 7	-77	8	809	None	
11	698	2014-09-18	753	198	Month 7	-73	8	800	None	
12	661	2014-08-12	717	176	Month 6	-76	5	756	None	

Q3. How many observations recorded a wind speed?

```
In [ ]: # query the collection for number of observations where wind speed is not "None"
data = collection.count_documents({"wind_speed": {"$ne": None}})

# print the number of observations
print("There are " + str(data) + " observations where wind speed was recorded.")
```

There are 0 observations where wind speed was recorded.

Q4. How many observations are recorded for each Martian month?

```
In [ ]: # query the collection for number of observations recorded per season
data = collection.aggregate([
    {"$group": {"_id": "$season", "count": {"$sum": 1}}},
    {"$sort": {"count": -1}}
])

# create a dataframe from the data
```

```
df = pd.DataFrame(data)

# rename the _id column to Martian month
df.rename(columns={"_id": "Martian month"}, inplace=True)

# show df
df.head(100)
```

Out[]:

	Martian month	count
0	Month 4	318
1	Month 3	308
2	Month 5	286
3	Month 6	266
4	Month 2	255
5	Month 11	245
6	Month 1	243
7	Month 7	239
8	Month 8	232
9	Month 12	224
10	Month 10	224
11	Month 9	196

Q5. What is the average temperature for each observation?

```
In [ ]: # avg_temp that is the average of the max and min temperature; include terrestrial
data = collection.aggregate([
    {"$project": {"terrestrial_date": 1, "season": 1, "min_temp": 1, "max_temp": 1},
    {"$sort": {"terrestrial_date": -1}}
])

# create a dataframe from the data
df = pd.DataFrame(data).drop(columns=["_id"])

# show df
df.head(10)
```

Out[]:

	terrestrial_date	season	min_temp	max_temp	avg_temp
0	2022-11-25	Month 12	-72.0	-44.0	-58.0
1	2022-11-24	Month 12	-72.0	-9.0	-40.5
2	2022-11-23	Month 12	-73.0	-8.0	-40.5
3	2022-11-22	Month 12	-72.0	-10.0	-41.0
4	2022-11-21	Month 12	-72.0	-12.0	-42.0
5	2022-11-20	Month 12	-71.0	-16.0	-43.5
6	2022-11-19	Month 12	-71.0	-6.0	-38.5
7	2022-11-18	Month 12	-70.0	-8.0	-39.0
8	2022-11-17	Month 12	-70.0	-7.0	-38.5
9	2022-11-16	Month 12	-70.0	-8.0	-39.0

Q6. What is the average temperature for each Martian month?

```
In [ ]: # query the collection for the average temperature per season
data = collection.aggregate([
    {"$project": {"season": 1, "avg_temp": {"$avg": ["$min_temp", "$max_temp"]}},
    {"$group": {"_id": "$season", "avg_temp": {"$avg": "$avg_temp"}}},
    {"$sort": {"avg_temp": -1}}
])

# create a dataframe from the data
df = pd.DataFrame(data)

# rename the _id column to Martian month
df.rename(columns={"_id": "Martian month"}, inplace=True)

# show df
df.head(100)
```

Out[]:

	Martian month	avg_temp
0	Month 8	-35.530172
1	Month 9	-36.645408
2	Month 7	-36.736402
3	Month 10	-37.103139
4	Month 11	-38.523013
5	Month 6	-39.830827
6	Month 12	-41.810268
7	Month 5	-46.117133
8	Month 1	-46.300412
9	Month 2	-51.058824
10	Month 4	-53.199686
11	Month 3	-54.663961

Q7. What is the average temperature during significant solar events (autumnal equinox, winter solstice, spring equinox, and summer solstice)?

```
In [ ]: # query collection for average temperature where ls is 0, 90, 180, or 270
data = collection.aggregate([
    {"$match": {"ls": {"$in": [0, 90, 180, 270]}},
    {"$project": {"ls": 1, "avg_temp": {"$avg": ["$min_temp", "$max_temp"]}},
    {"$group": {"_id": "$ls", "avg_temp": {"$avg": "$avg_temp"}},
    {"$sort": {"_id": 1}}
])

# create a dataframe from the data
df = pd.DataFrame(data)

# rename the _id column to ls
df.rename(columns={"_id": "ls (solar longitude)"}, inplace=True)

# new column for name of event
df["event"] = ["Autumnal Equinox", "Winter Solstice", "Spring Equinox", "Summer Solstice"]

# reorder columns
df = df[["ls (solar longitude)", "event", "avg_temp"]]

# show df
df.head()
```

Out []:

	ls (solar longitude)	event	avg_temp
0	0	Autumnal Equinox	-43.375000
1	90	Winter Solstice	-55.454545
2	180	Spring Equinox	-36.000000
3	270	Summer Solstice	-36.428571

Q8. What weather conditions have been observed on Mars, and how many times did they occur?

```
In [ ]: # query collection for what values exist for atmospheric opacity and how many times
data = collection.aggregate([
    {"$group": {"_id": "$atmo_opacity", "count": {"$sum": 1}}},
    {"$sort": {"count": -1}}
])

# create a dataframe from the data
df = pd.DataFrame(data)

# rename the _id column to atmo_opacity
df.rename(columns={"_id": "atmo_opacity (weather condition)"}, inplace=True)

# show df
df.head()
```

Out []:

	atmo_opacity (weather condition)	count
0	Sunny	3033
1	--	3

Q9. What is the average pressure (and temperature) per Martian month?

```
In [ ]: # query collection for average pressure and temp per season
data = collection.aggregate([
    {"$project": {"season": 1, "pressure": 1, "avg_temp": {"$avg": ["$min_temp", "$max_temp"]}},
    {"$group": {"_id": "$season", "avg_pressure": {"$avg": "$pressure"}, "avg_temp": {"$avg": "$avg_temp"}},
    {"$sort": {"avg_pressure": -1}}
])

# create a dataframe from the data
df = pd.DataFrame(data)

# rename the _id column to Month
df.rename(columns={"_id": "Martian month"}, inplace=True)

# show df
df.head(100)
```

Out[]:

	Martian month	avg_pressure	avg_temp
0	Month 9	895.163265	-36.645408
1	Month 2	876.556863	-51.058824
2	Month 10	870.206278	-37.103139
3	Month 3	866.931818	-54.663961
4	Month 8	854.340517	-35.530172
5	Month 1	847.127572	-46.300412
6	Month 11	835.708333	-38.523013
7	Month 12	824.415179	-41.810268
8	Month 4	798.242138	-53.199686
9	Month 7	777.782427	-36.736402
10	Month 5	736.388112	-46.117133
11	Month 6	728.823308	-39.830827

Q10. What is the highest difference in pressure recorded on Mars?

```
In [ ]: # query collection for highest and lowest pressure and the difference between 1
data = collection.aggregate([
    {"$project": {"pressure": 1}},
    {"$group": {"_id": None, "max_pressure": {"$max": "$pressure"}, "min_pressu
    {"$addFields": {"largest_pressure_diff": {"$subtract": ["$max_pressure", "$s
])

# create a dataframe from the data
df = pd.DataFrame(data).drop(columns=["_id"])

# show df
print(df.head())

# print the largest pressure difference
print("\nThe largest difference in atmospheric pressure recorded on Mars is " +
```

	max_pressure	min_pressure	largest_pressure_diff
0	925	702	223

The largest difference in atmospheric pressure recorded on Mars is 223 Pa.

Visualization

Plot for Q9

```
In [ ]: # query Q9
data = collection.aggregate([
    {"$project": {"season": 1, "pressure": 1, "avg_temp": {"$avg": ["$min_temp",
    {"$group": {"_id": "$season", "avg_pressure": {"$avg": "$pressure"}, "avg_t
```



```

    {"$sort": {"avg_pressure": -1}}
  ])

# create a dataframe from the data
df = pd.DataFrame(data)

# rename the _id column to Month
df.rename(columns={"_id": "Martian month"}, inplace=True)

# create column for month number based on part Martian month after "Month "
df["month_num"] = df["Martian month"].str.split(" ").str[1]

# convert month_num to int and sort by month_num
df["month_num"] = df["month_num"].astype(int)
df.sort_values(by="month_num", inplace=True)

# show df
df.head(100)

```

Out []:

	Martian month	avg_pressure	avg_temp	month_num
5	Month 1	847.127572	-46.300412	1
1	Month 2	876.556863	-51.058824	2
3	Month 3	866.931818	-54.663961	3
8	Month 4	798.242138	-53.199686	4
10	Month 5	736.388112	-46.117133	5
11	Month 6	728.823308	-39.830827	6
9	Month 7	777.782427	-36.736402	7
4	Month 8	854.340517	-35.530172	8
0	Month 9	895.163265	-36.645408	9
2	Month 10	870.206278	-37.103139	10
6	Month 11	835.708333	-38.523013	11
7	Month 12	824.415179	-41.810268	12

In []:

```

# make subplots
fig = make_subplots(rows=2, cols=1, shared_xaxes=True, vertical_spacing=0.05,

# add traces
fig.add_trace(go.Scatter(x=df["Martian month"], y=df["avg_pressure"], name="Average Pressure"))
fig.add_trace(go.Scatter(x=df["Martian month"], y=df["avg_temp"], name="Average Temperature"))

# update layout
fig.update_layout(height=600, width=800, title_text="Average Pressure and Temperature")

# show figure
fig.show()

```

Interpretation

This plot shows the average pressure and temperature over each Martian month. The correlation between the Martian month and pressure demonstrates a cyclical pattern. The correlation between the Martian month and temperature demonstrates similar cyclical patterns, but with a larger "wavelength" than the pressure. The temperature decreases and increases more slowly throughout a Martian year than the pressure. Earlier in the Martian year, the temperature and pressure are negatively correlated, however, later in the Martian year, the temperature and pressure become positively correlated, as pressure increases more rapidly than temperature. This probably has to do with how the data is collected by the Curiosity rover. The rover is likely to be more active during the warmer months, and less active during the colder months. This would explain why the pressure and temperature are more correlated later in the Martian year.