Drought Predictions

Carolyn Chu, Romario Leal, Amy Paschal, David Rodgers, Austin McClain

Data Set

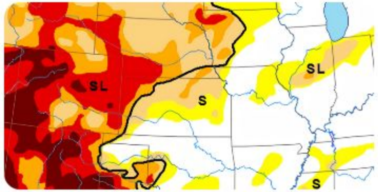
https://www.kaggle.com/datasets/cdminix/us-drought-meteorological-data

CHRISTOPH MINIXHOFER · UPDATED 2 YEARS AGO

181 | New Notebook | Download (945 MB)

Predict Droughts using Weather & Soil Data

Predicting continental US drought levels using meteorological & soil data.



Data Card | Code (7) | Discussion (5)

About Dataset

Update (14/12/21): Kaggle Tasks are being deprecated, so I moved the current results on this dataset here:

User	*Model/Notebook*	Macro F1 Mean	MAE Mean

Usability ⓘ
10.00

License
CC0: Public Domain

Expected update frequency

Two Data Sets

Soil Data.csv

Shape: 3,109 rows, 32 columns

Description of the measurement sites

Time Series Data.csv

Shape: Over 23 million rows, 21 columns

Pre-Split between Training, Validation, and Test

Data on the weather conditions at each site over 20 years, and a drought score

Soil_data.csv

(Used FIPS, latitude and longitude in map visualization)

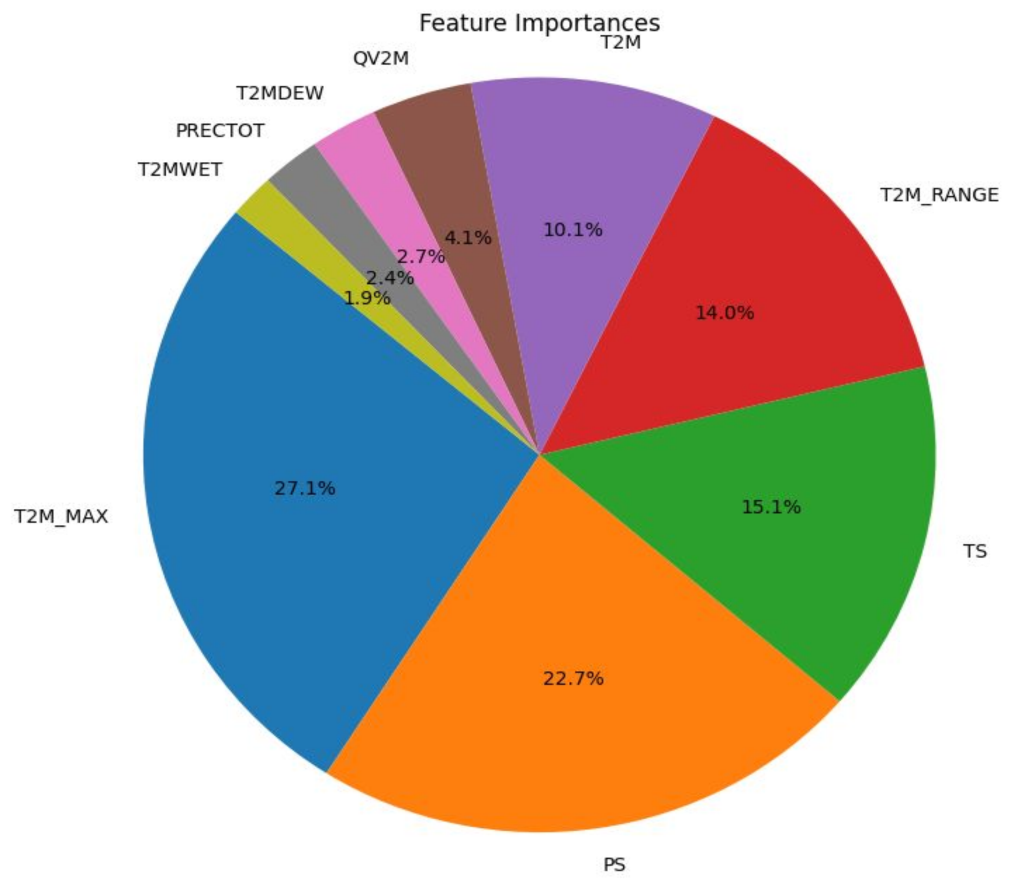
- Fips: US county FIPS code. (Geocode for County)
- Lat
- Lon
- Elevation: median elevation (meters)
- Slope1: 0 % ≤ slope ≤ 0.5 %
- Slope2: 0.5 % ≤ slope ≤ 2 %
- Slope3: 2 % ≤ slope ≤ 5 %
- Slope4: 5 % ≤ slope ≤ 10 %
- Slope5: 10 % ≤ slope ≤ 15 %
- Slope6: 15 % ≤ slope ≤ 30 %
- Slope7: 30 % ≤ slope ≤ 45 %
- aspectN: North: 0˚< aspect ≤45˚ or 315˚< aspect ≤360˚
- espectE: East: 45˚ < aspect ≤ 135˚
- aspectS: South: 135˚ < aspect ≤ 225˚
- aspectW: West: 225˚ < aspect ≤ 315˚
- aspectUnknown: Undefined: Slope aspect undefined; this value is used for grids where slope gradient is undefined or slope gradient is less than 2%.
- WAT_LAND: mapped water bodies
- NVG_LAND: barren/very sparsely vegetated land
- URB_LAND: built-up land (residential and infrastructure)
- GRS_LAND: grass/scrub/woodland
- FOR_LAND: forrest land, calibrated to FRA2000 land statistics
- CULTRF_LAND
- CULTIR_LAND: irrigated cultivated land, according to GMIA 4.0
- CULT_LAND: total cultivated land
- SQ1: Nutrient availability
- SQ2: Nutrient retention capacity
- SQ3: Rooting conditions
- SQ4: Oxygen availability to roots
- SQ5: excess salts
- SQ6: toxicity
- SQ7: Workability

Time Series Data

(Used in Machine Learning Model)

- Fips (County Number)
- Date
- PRECTOT: Precipitation (mm day-1)
- PS: Surface Pressure (kPa)
- QV2M: Specific Humidity at 2 Meters (g/kg)
- T2M: Temperature at 2 Meters ©
- T2MDEW: Dew/Frost Point at 2 Meters ©
- T2MWET: Wet Bulb Temperature at 2 Meters
- T2M_MAX: Maximum Temperature at 2 Meters
- T2M_MIN: Minimum Temperature at 2 Meters
- T2M_RANGE: Temperature Range at 2 Meters
- TS: Earth Skin Temperature ©
- WS10M: Wind Speed at 10 Meters (m/s)
- WS10M_MAX: Maximum Wind Speed at 10 Meters (m/s)
- WS10M_MIN: Minimum Wind Speed at 10 Meters (m/s)
- WS10M_RANGE: Wind Speed Range at 10 Meters (m/s)
- WS50M: Wind Speed at 50 Meters (m/s)
- WS50M_MAX: Maximum Wind Speed at 50 Meters (m/s)
- WS50M_MIN: Minimum Wind Speed at 50 Meters (m/s)
- WS50M_RANGE: Wind Speed Range at 50 Meters (m/s)
- Score: Measure of drought ranging from 0 (no drought) to 5 (D4, see description).

Important Features



- T2M_MAX: Maximum Temperature at 2 Meters
- T2MWET: Wet Bulb Temperature at 2 Meters
- T2MDEW: Dew/Frost Point at 2 Meters
- T2M: Temperature at 2 Meters
- T2M_RANGE: Temperature Range at 2 Meters
- TS: Earth Skin Temperature
- PRECTOT: Precipitation (mm day-1)
- QV2M: Specific Humidity at 2 Meters (g/kg)
- PS: Surface Pressure (kPa)

Target: Drought Score

Image source: https://droughtmonitor.unl.edu

Category	Description	Possible Impacts
D0	Abnormally Dry	Going into drought: ▪ short-term dryness slowing planting, growth of crops or pastures Coming out of drought: ▪ some lingering water deficits ▪ pastures or crops not fully recovered
D1	Moderate Drought	▪ Some damage to crops, pastures ▪ Streams, reservoirs, or wells low, some water shortages developing or imminent ▪ Voluntary water-use restrictions requested
D2	Severe Drought	▪ Crop or pasture losses likely ▪ Water shortages common ▪ Water restrictions imposed
D3	Extreme Drought	▪ Major crop/pasture losses ▪ Widespread water shortages or restrictions
D4	Exceptional Drought	▪ Exceptional and widespread crop/pasture losses ▪ Shortages of water in reservoirs, streams, and wells creating water emergencies

(image source: https://droughtmonitor.unl.edu)

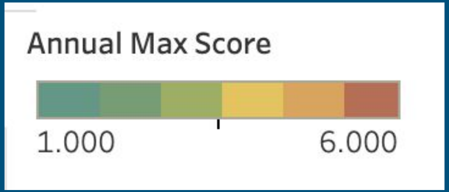
Tools Used

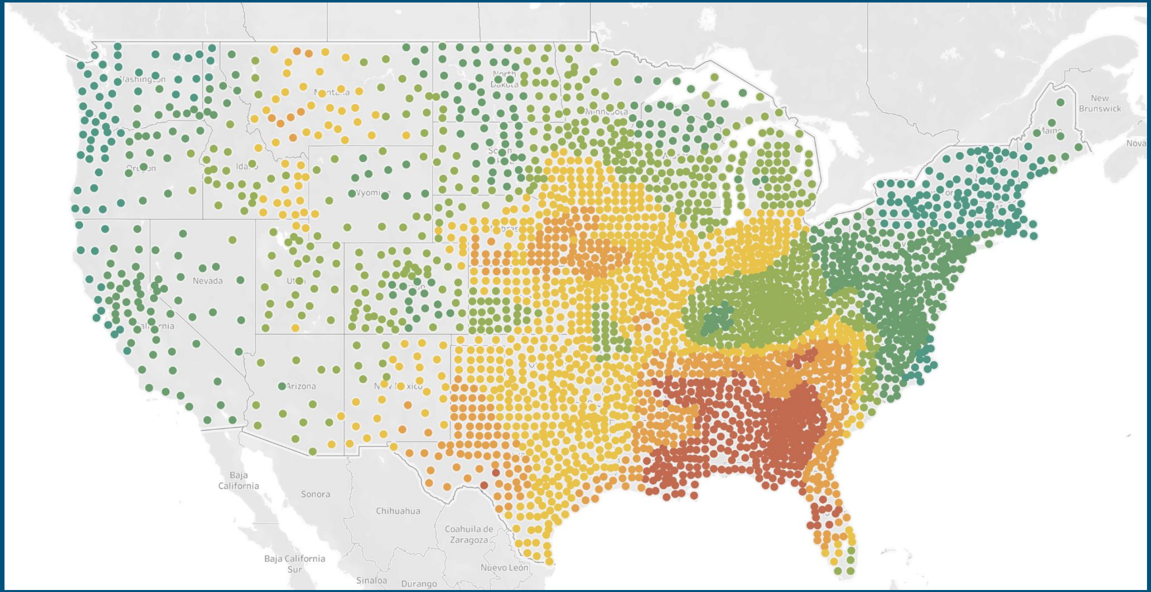
- Colab Notebook, PySpark - Because of the large dataset
- PySpark.ML
- PySpark.SQL
- Tableau - Used to create interactive map
- Google Drive - To host data
- Pandas
- Matplotlib.pyplot
- GeoPy

Interactive Map of Drought Scores

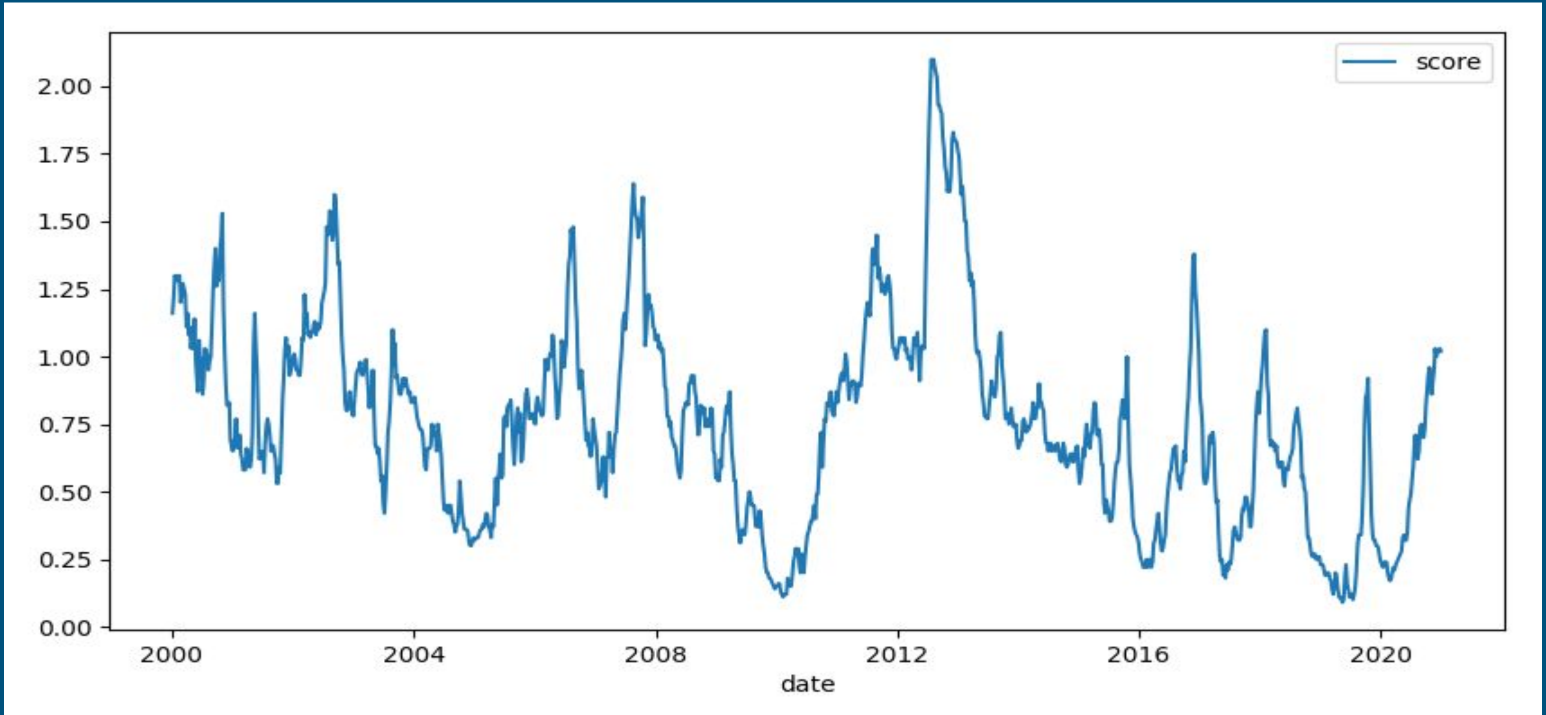
Visualization of the drought scores measured at each location over the 2 decades of samples

https://public.tableau.com/app/profile/amy.paschal/viz/DroughtDataAnnual/AnnualDroughtScores?publish=yes

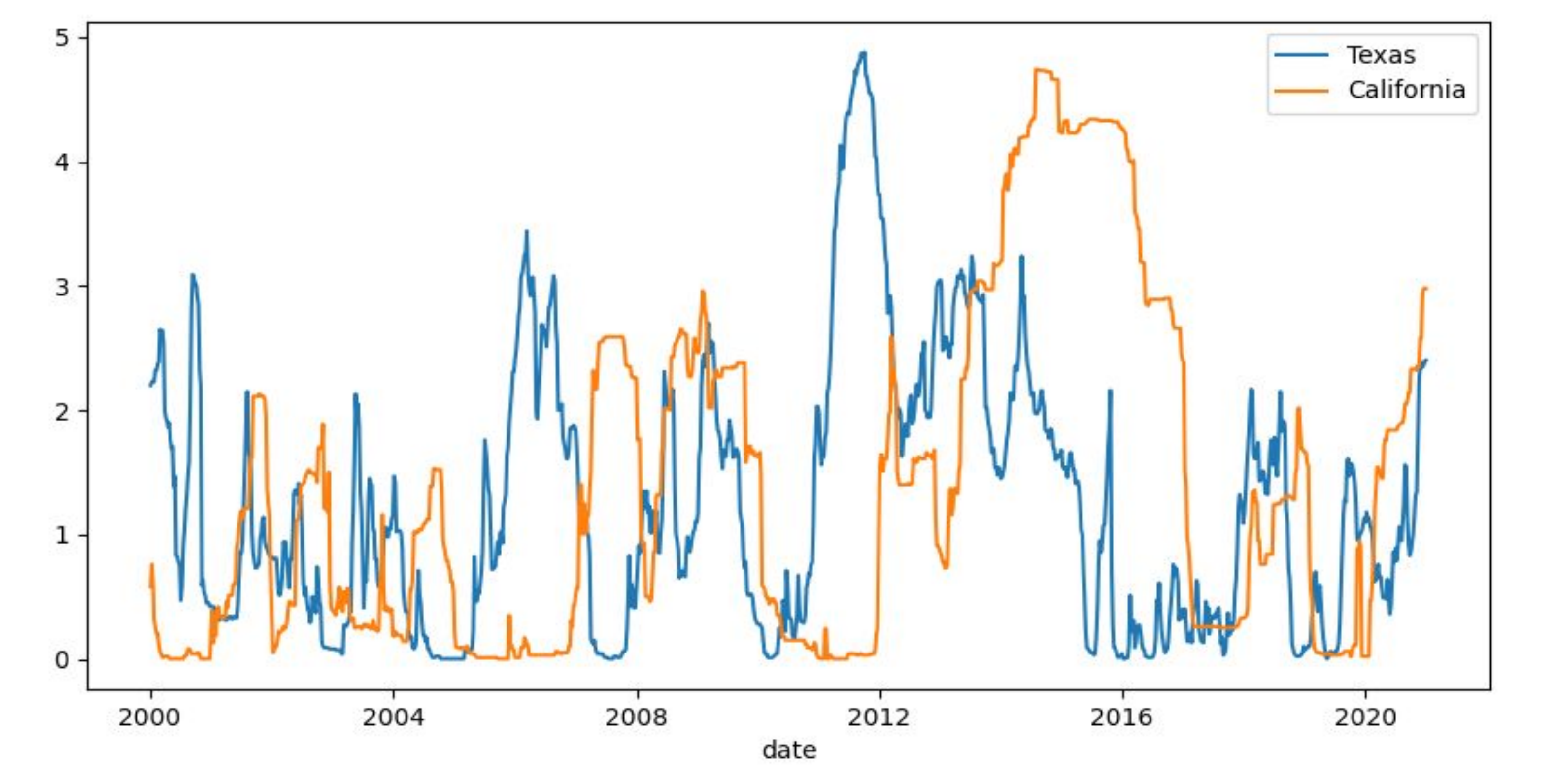




Time Series of Droughts



State Comparison of Droughts

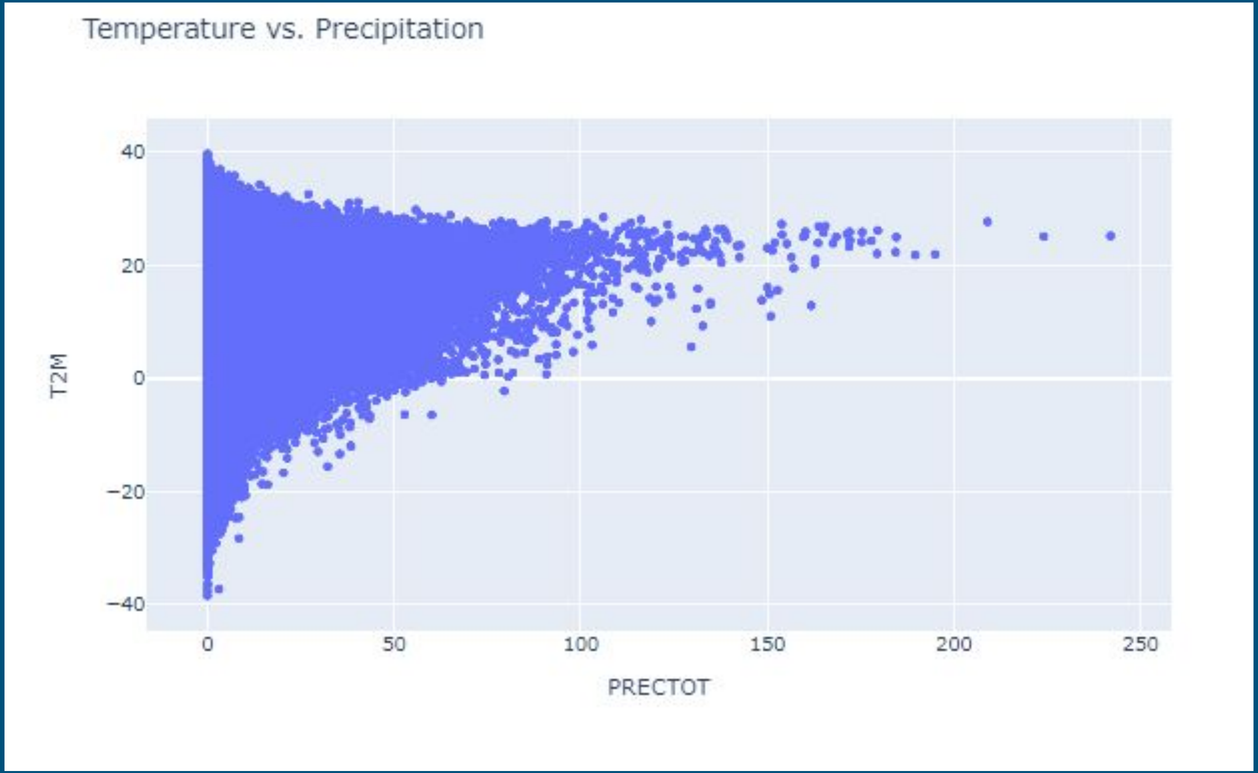


Correlation Within the Data

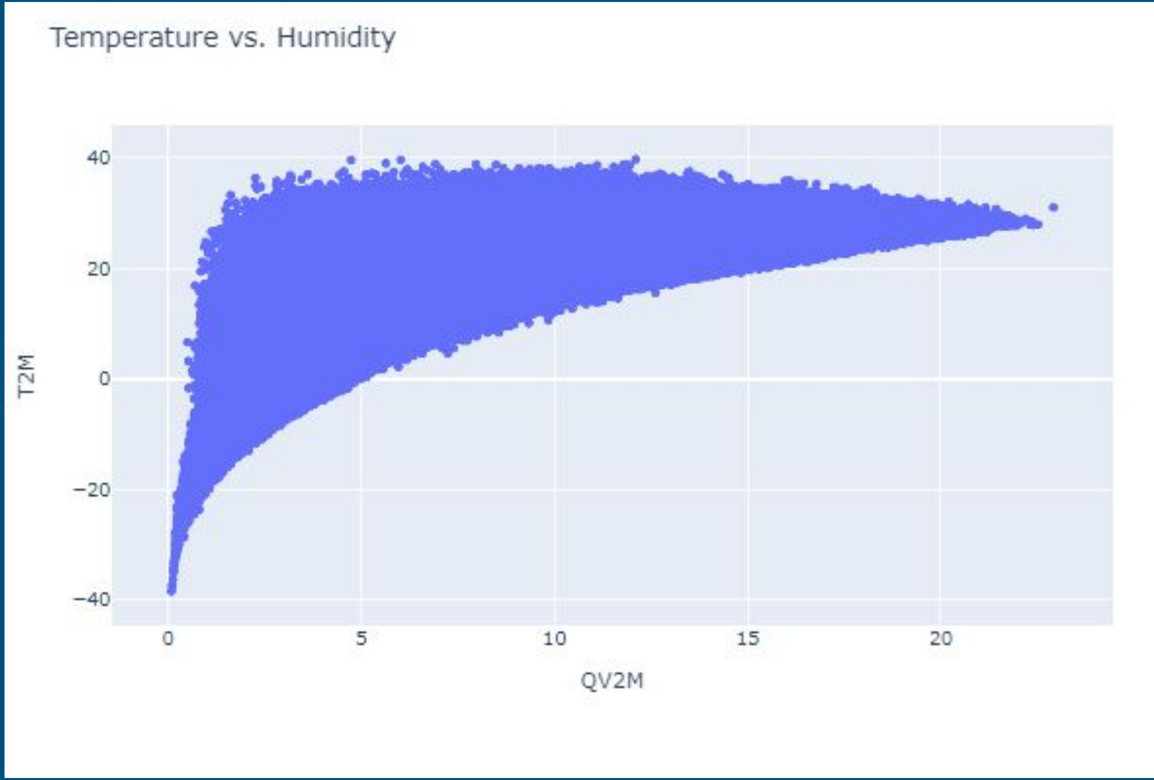
```
The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
```

	fips	PRECTOT	PS	QV2M	T2M	T2MDEW	T2MWET	T2M_MAX	T2M_MIN	T2M_RANGE	TS	WS10M	WS10M_MAX	WS10M_MIN	WS10M_RANGE	WS50M	WS50M_MAX	WS50M_MIN	WS50M_RANGE	score
fips	1.000000	-0.017850	-0.036142	-0.056147	-0.049778	-0.051682	-0.051917	-0.051610	-0.047799	-0.022288	-0.045461	0.046921	0.041874	0.035454	0.031910	0.051558	0.053576	0.037171	0.034296	-0.027649
PRECTOT	-0.017850	1.000000	0.069296	0.246240	0.100873	0.232173	0.232087	0.036727	0.150230	-0.292497	0.097215	0.061849	0.072254	0.032902	0.074648	0.077991	0.087799	0.062441	0.054711	-0.057248
PS	-0.036142	0.069296	1.000000	0.279132	0.163906	0.339142	0.339296	0.112127	0.207447	-0.226464	0.163713	-0.076733	-0.131601	0.026697	-0.195957	-0.039712	-0.087153	0.039040	-0.152824	-0.193770
QV2M	-0.056147	0.246240	0.279132	1.000000	0.874161	0.959216	0.960844	0.809535	0.908280	-0.071035	0.866537	-0.224791	-0.252893	-0.116956	-0.260021	-0.208001	-0.247849	-0.091739	-0.236876	-0.059532
T2M	-0.049778	0.100873	0.163906	0.874161	1.000000	0.915951	0.916536	0.983373	0.981873	0.235958	0.997493	-0.210898	-0.219989	-0.136063	-0.202416	-0.198665	-0.207992	-0.123925	-0.153011	0.077679
T2MDEW	-0.051682	0.232173	0.339142	0.959216	0.915951	1.000000	0.999783	0.858401	0.940536	-0.015436	0.907692	-0.237314	-0.264393	-0.123948	-0.270686	-0.205759	-0.242316	-0.092143	-0.229197	-0.066244
T2MWET	-0.051917	0.232087	0.339296	0.960844	0.916536	0.999783	1.000000	0.858840	0.941400	-0.016458	0.908295	-0.237426	-0.264564	-0.124051	-0.270846	-0.206246	-0.242842	-0.092488	-0.229552	-0.064836
T2M_MAX	-0.051610	0.036727	0.112127	0.809535	0.983373	0.858401	0.858840	1.000000	0.938408	0.399326	0.980104	-0.219668	-0.221403	-0.152359	-0.193035	-0.200980	-0.197687	-0.144130	-0.119738	0.116188
T2M_MIN	-0.047799	0.150230	0.207447	0.908280	0.981873	0.940536	0.941400	0.938408	1.000000	0.057946	0.979428	-0.209927	-0.226195	-0.123666	-0.219371	-0.203867	-0.227677	-0.108317	-0.194149	0.046264
T2M_RANGE	-0.022288	-0.292497	-0.226464	-0.071035	0.235958	-0.015436	-0.016458	0.399326	0.057946	1.000000	0.233000	-0.077668	-0.039515	-0.112074	0.024341	-0.039753	0.032938	-0.129024	0.169191	0.212859
TS	-0.045461	0.097215	0.163713	0.866537	0.997493	0.907692	0.908295	0.980104	0.979428	0.233000	1.000000	-0.193360	-0.202963	-0.121362	-0.189638	-0.186798	-0.195444	-0.114239	-0.145933	0.084699
WS10M	0.046921	0.061849	-0.076733	-0.224791	-0.210898	-0.237314	-0.237426	-0.219668	-0.209927	-0.077668	-0.193360	1.000000	0.952973	0.836162	0.705937	0.966765	0.911773	0.799846	0.420517	0.023874
WS10M_MAX	0.041874	0.072254	-0.131601	-0.252893	-0.219989	-0.264393	-0.264564	-0.221403	-0.226195	-0.039515	-0.202963	0.952973	1.000000	0.694954	0.867155	0.911357	0.947880	0.666507	0.598089	0.044131
WS10M_MIN	0.035454	0.032902	0.026697	-0.116956	-0.136063	-0.123948	-0.124051	-0.152359	-0.123666	-0.112074	-0.121362	0.836162	0.694954	1.000000	0.244522	0.841849	0.674361	0.944219	-0.032867	-0.012678
WS10M_RANGE	0.031910	0.074648	-0.195957	-0.260021	-0.202416	-0.270686	-0.270846	-0.193035	-0.219371	0.024341	-0.189638	0.705937	0.867155	0.244522	1.000000	0.645878	0.811137	0.244792	0.829289	0.067901
WS50M	0.051558	0.077991	-0.039712	-0.208001	-0.198665	-0.205759	-0.206246	-0.200980	-0.203867	-0.039753	-0.186798	0.966765	0.911357	0.841849	0.645878	1.000000	0.920380	0.851838	0.381149	0.001611
WS50M_MAX	0.053576	0.087799	-0.087153	-0.247849	-0.207992	-0.242316	-0.242842	-0.197687	-0.227677	0.032938	-0.195444	0.911773	0.947880	0.674361	0.811137	0.920380	1.000000	0.656003	0.676956	0.022690
WS50M_MIN	0.037171	0.062441	0.039040	-0.091739	-0.123925	-0.092143	-0.092488	-0.144130	-0.108317	-0.129024	-0.114239	0.799846	0.666507	0.944219	0.244792	0.851838	0.656003	1.000000	-0.111432	-0.028690
WS50M_RANGE	0.034296	0.054711	-0.152824	-0.236876	-0.153011	-0.229197	-0.229552	-0.119738	-0.194149	0.169191	-0.145933	0.420517	0.598089	-0.032867	0.829289	0.381149	0.676956	-0.111432	1.000000	0.057221
score	-0.027649	-0.057248	-0.193770	-0.059532	0.077679	-0.066244	-0.064836	0.116188	0.046264	0.212859	0.084699	0.023874	0.044131	-0.012678	0.067901	0.001611	0.022690	-0.028690	0.057221	1.000000

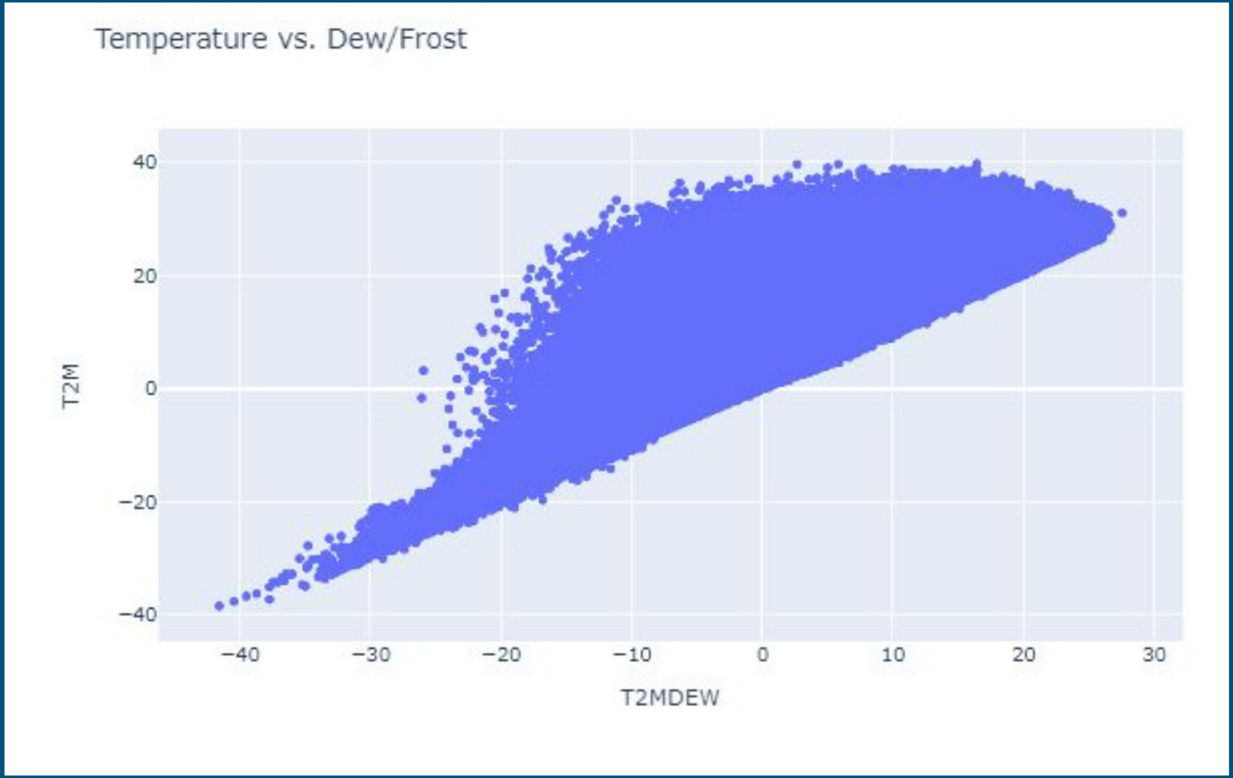
Temperature vs Precipitation



Temperature vs Humidity

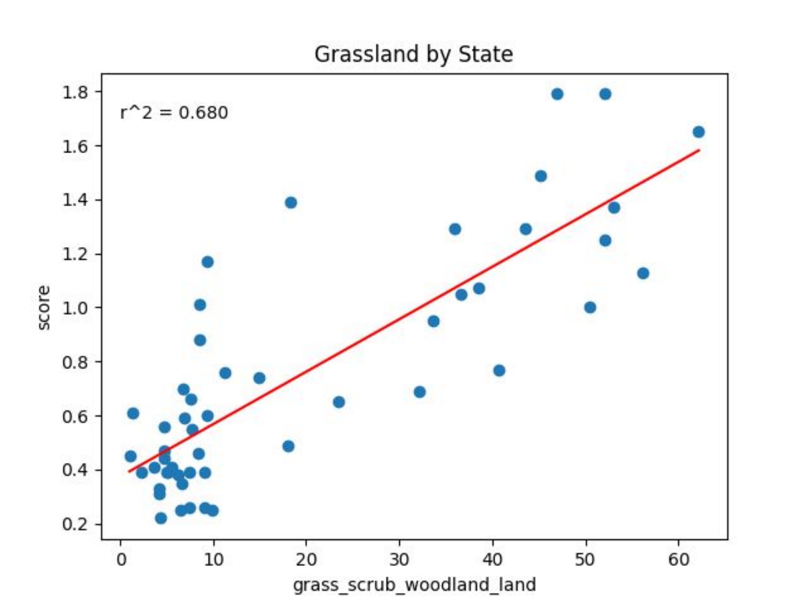


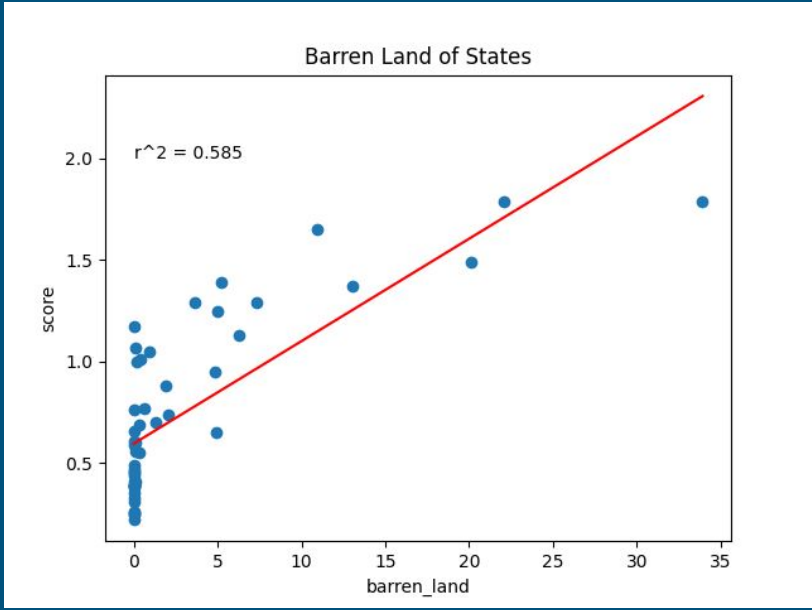
Temperature vs Dew/Frost



Linear Regressions with the Soil Data

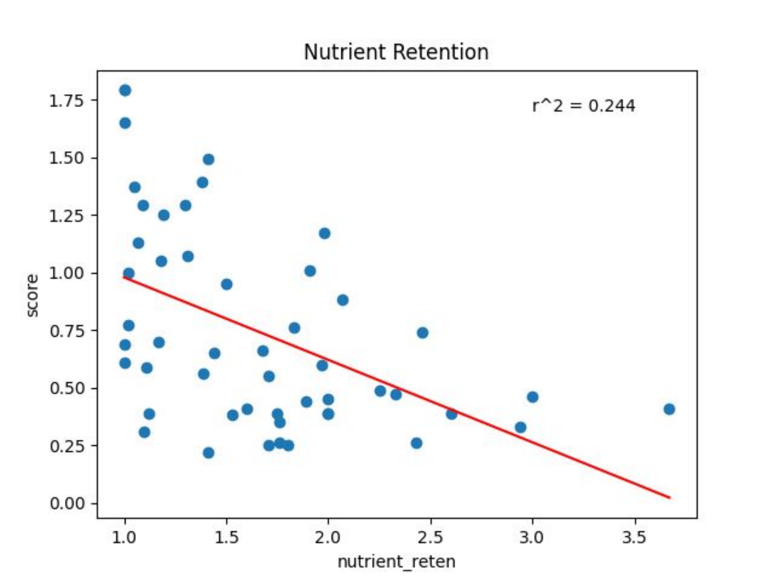
Type of Land

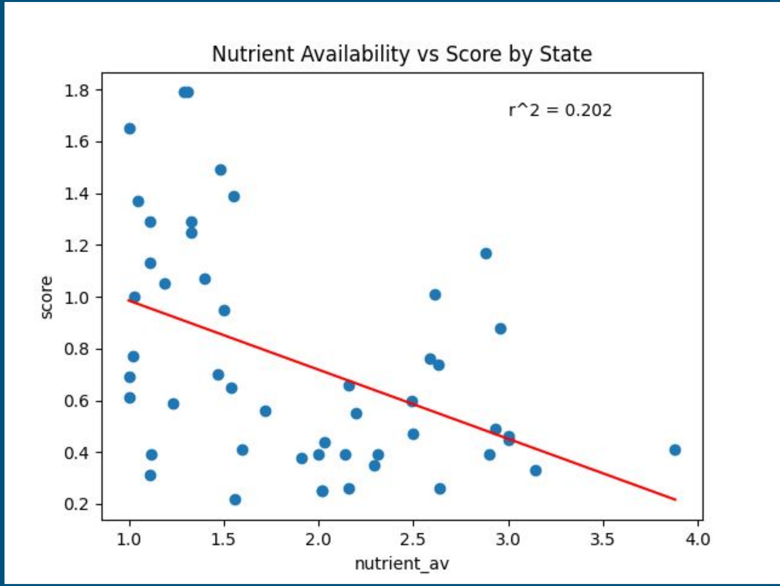




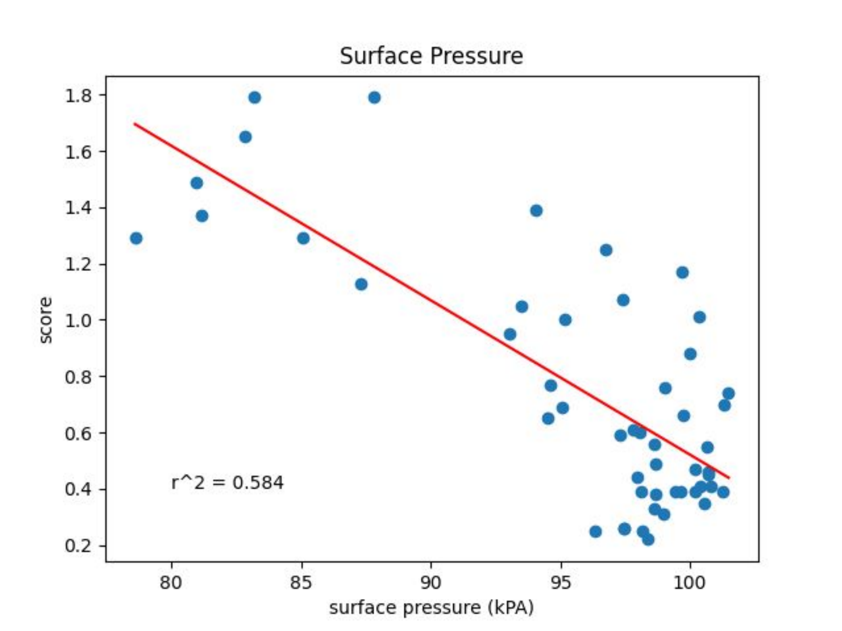
Linear Regressions with the Soil Data

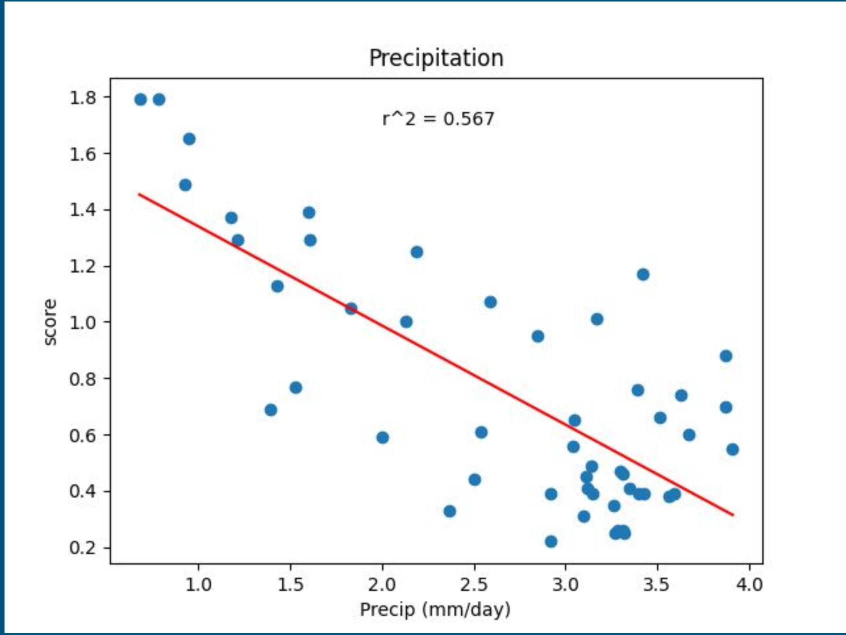
Soil Features



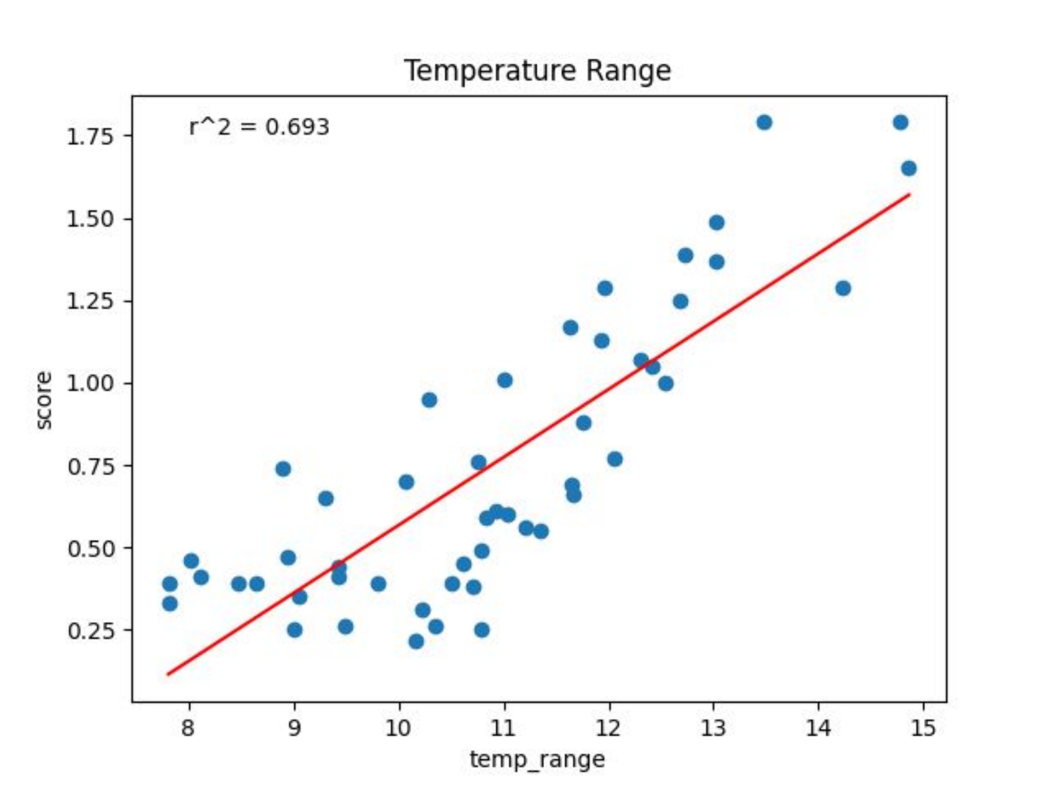


Linear Regressions with the Time Series Data





Linear Regressions with the Time Series Data



Machine Learning Model

PySpark - Vector Assembler to Create Dense Vector Column

PRECTOT	PS	QV2M	T2M	T2MDEW	WS50M_MAX	WS50M_MIN	WS50M_RANGE	score	features	scaled_features
15.95	100.29	6.42	11.4	6.09	9.31	3.74	5.58	1	[15.95,100.29,6.4...	[2.11848627313622...
1.33	100.4	6.63	11.48	7.84	6.38	1.71	4.67	2	[1.33,100.4,6.63,...	[-0.221616093985O...
1.11	100.39	9.53	14.28	13.26	6.4	3.84	2.55	2	[1.11,100.39,9.53...	[-0.2568296726968...
0.0	100.11	2.05	-0.78	-7.93	8.03	3.96	4.07	2	[0.0,100.11,2.05,...	[-0.4344981834700...
0.0	101.0	3.36	2.06	-1.73	6.38	1.27	5.11	1	[0.0,101.0,3.36,2...	[-0.4344981834700...
0.0	101.04	5.01	8.88	3.9	5.33	0.69	4.63	1	[0.0,101.04,5.01,...	[-0.4344981834700...
0.0	100.68	5.84	10.94	5.84	7.63	2.57	5.07	1	[0.0,100.68,5.84,...	[-0.4344981834700...
0.36	101.34	5.03	10.02	3.99	7.13	2.67	4.46	1	[0.36,101.34,5.03...	[-0.3768759637598...
0.01	100.71	6.52	12.61	7.49	7.64	2.1	5.54	1	[0.01,100.71,6.52...	[-0.4328975662558...
0.0	100.79	7.48	16.81	9.47	6.92	1.84	5.07	1	[0.0,100.79,7.48,...	[-0.4344981834700...
0.02	100.8	5.34	11.24	4.62	7.29	1.99	5.3	1	[0.02,100.8,5.34,...	[-0.4312969490416...
0.17	100.66	6.89	12.81	8.32	5.55	1.55	4.0	2	[0.17,100.66,6.89...	[-0.4072876908290...
0.0	99.07	7.46	14.58	9.1	8.68	3.52	5.16	1	[0.0,99.07,7.46,1...	[-0.4344981834700...
1.26	100.32	5.52	10.67	4.67	9.16	3.79	5.37	1	[1.26,100.32,5.52...	[-0.2328204144842...
0.36	100.77	9.52	16.14	13.19	4.87	2.73	2.14	1	[0.36,100.77,9.52...	[-0.3768759637598...
0.0	100.1	7.93	14.82	10.43	6.13	3.87	2.26	1	[0.0,100.1,7.93,1...	[-0.4344981834700...
0.04	99.77	7.66	14.93	9.93	7.01	4.58	2.43	1	[0.04,99.77,7.66,...	[-0.4280957146133...
0.31	100.28	11.93	21.87	16.72						
0.04	99.84	10.61	23.46	14.68						

PySpark Machine Learning Algorithms Attempted

Multilayer Perceptron Classifier	Linear Regression Model	Decision Tree Regressor	Random Forest Regressor	Random Forest Classifier
- Rounded Scores to create five bins - Kept giving errors during evaluation	- Root Mean Squared Error (RMSE) of 0.5846… - Didn't produce feature importance	RMSE = 0.590201 Mean Squared Error (MSE) = 0.348337 Mean Absolute Error (MAE) = 0.38777 R Squared (R2) = -0.00464572	Similar results to Decision Tree Regressor	- Rounded scores - Produced Feature Importance - Accuracy of 0.7591