

# Predictive Modeling of Effect of Environmental Conditions on Road Construction Projects in USA

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-jakt@534307f](#) on October 28, 2022.

## Authors

---

- **Amirthavarshini Muraleetharan**

 [677-010-487](#) ·  [amirthavarshini246](#)

Department of CEE, University of Illinois Urbana Champaign

- **Thomas Ngare**

 [652-601-317](#) ·  [thomasNg](#)

Department of CEE, University of Illinois Urbana Champaign

- **Kapil Shah**

 [668-376-620](#) ·  [kapilrs2](#)

Department of CEE, University of Illinois Urbana Champaign

# Abstract

---

## 1.0 Introduction

Table 1: Description of Undertaking Dataset

Features	Description
ID	Unique identifier of construction record
Severity	Shows the severity of the construction
Start and End Time	Shows the start time of construction
End Time	Shows the end time of construction
Latitude and Longitude	Shows the GPS coordinates
Distance	The length of the road extent affected by the construction
Street Details	Shows the street number, name and right/left side in address field
Address Details	Shows the city, county, state, country and zip code in address field
Time zone	Shows time zone based on the location of the construction event
Weather	Shows the time stamp of weather observation record
Temperature, Wind, Humidity, and Pressure	Shows the temperature, wind chill, humidity, and pressure
Visibility	Shows visibility
Wind Direction and Speed	Shows wind conditions
Precipitation and Weather condition	Shows precipitation and weather condition
Amenity	An annotation which indicates presence of amenity in a nearby location
Bump and Crossing	Annotations which indicate presence of speed bump or hump and crossings
Give way, Junction, railway	Annotations which indicate presence of give way, junction and railway
Exit, Roundabout, Station, Stop	Annotation which indicates presence of no exit, railway, roundabout, and station
Traffic Details	Annotations which indicate traffic calming, signal, turning loop
Light Details	Annotations which indicate sunrise, sunset, civil twilight, nautical twilight, astronomical twilight

## 2.0 Data Wrangling and Exploratory Data Analysis

### 2.1.1 Data Cleaning

Table 2: Dataset of road construction projects in USA

Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Pressure(in)	Visibility(mi)	Wind_Direction	Wind_Speed(mph)	Precipitation(in)
String31	"9/29/2020 11:53"	32.8384	-93.1524	32.8507	-93.1644	1.1035	29.72	10.0	"S"	3.0	0.0
"4/5/2019 16:00"	"11/12/2021 8:22"	30.2213	-92.0086	30.2166	-92.0038	0.433173	30.09	3.0	"CALM"	0.0	0.0
"11/12/2021 7:59"	"10/12/2021 9:18"	39.6532	-104.91	39.6531	-104.914	0.192266	24.09	10.0	"WSW"	5.0	0.0
"10/12/2021 7:17"	"2/17/2021 23:59"	33.9615	-118.029	33.9619	-118.029	0.0321121	29.92	9.0	"CALM"	0.0	0.0
"2/10/2021 2:46"	"12/6/2021 9:53"	27.8957	-82.7872	27.8946	-82.7853	0.141399	30.12	1.0	"SE"	5.0	0.0
"12/6/2021 7:50"	"9/30/2021 3:40"	41.977	-87.9398	41.9759	-87.9398	0.0732397	29.39	10.0	"CALM"	0.0	0.0
"9/30/2021 1:39"	"9/16/2021 15:40"	38.9209	-77.0365	38.9333	-77.0365	0.853305	30.09	10.0	"N"	20.0	0.01
"9/16/2021 15:08"	"12/1/2021 10:56"	40.8683	-73.9191	40.867	-73.9231	0.226995	30.15	10.0	"W"	16.0	0.0
"12/1/2021 8:51"	"7/31/2021 13:06"	33.4831	-112.074	33.4793	-112.074	0.258762	28.83	10.0	"WNW"	8.0	0.0
"7/31/2021 12:11"	"11/19/2021 16:27"	33.4949	-112.028	33.4954	-112.03	0.130887	28.81	10.0	"ESE"	6.0	0.0
"11/19/2021 14:06"											
"7/22/2021 10:45"	"7/22/2021 11:31"	31.7066	-92.2451	31.7057	-92.2348	0.607945	30.04	10.0	"W"	9.0	0.0

Table 3: A summary of pertinent information of road construction projects in USA

variable	mean	min	median	median	max	nmissing	eltype
1 :ID	nothing	"C-1"	nothing	1 nothing	"C-999999"	0	String15
2 :Severity	2.25385	1	2.0	2 2.0	4	0	Int64
3 :Start_Time	nothing	"1/1/2016 3:01"	nothing	3 nothing	"9/9/2021 9:59"	0	String31
4 :End_Time	nothing	"1/1/2017 0:45"	nothing	4 nothing	"9/9/2021 9:58"	0	String31
5 :Start_Lat	37.2934	21.3551	39.2578	5 39.2578	49.0008	0	Float64
6 :Start_Lng	-91.436	-157.939	-85.6933	6 -85.6933	-67.0758	0	Float64
7 :End_Lat	37.1675	21.3548	39.198	7 39.198	49.005	107566	Union{Missing, Float64}
8 :End_Lng	-91.3286	-157.94	-85.6538	8 -85.6538	-67.0747	107566	Union{Missing, Float64}
9 Symbol("Distance(mi)")	1.2595	0.0	0.325327	9 0.325327	424.374	0	Float64
10 :Description	nothing	" and will be enforced by local law en	nothing	10 n nothing	"working behind barrier walls. CHARLES	0	String
: more				: more			
47 :AstronomicalTwilight	nothing	"Day"	nothing	47 nothing	"Night"	99	Union{Missing, String7}

2.2 Wrangling and feature derivation

Table 4: Derivation of project duration from raw data

Data before wrangling			Data after wrangling		
Start_Time	End_Time	Distance(mi)	Start_Date	End_Date	Project_Duration_Days
"4/5/2019 16:00"	"9/29/2020 11:53"	1.1035	2019-04-05	2020-09-29	544
"11/12/2021 7:59"	"11/12/2021 8:22"	0.433173	2021-11-12	2021-11-12	1
"10/12/2021 7:17"	"10/12/2021 9:18"	0.192266	2021-10-12	2021-10-12	1
"2/10/2021 2:46"	"2/17/2021 23:59"	0.0321121	2021-02-10	2021-02-17	8
"12/6/2021 7:50"	"12/6/2021 9:53"	0.141399	2021-12-06	2021-12-06	1
"9/30/2021 1:39"	"9/30/2021 3:40"	0.0732397	2021-09-30	2021-09-30	1
"9/16/2021 15:08"	"9/16/2021 15:40"	0.853305	2021-09-16	2021-09-16	1
"12/1/2021 8:51"	"12/1/2021 10:56"	0.226995	2021-12-01	2021-12-01	1
"7/31/2021 12:11"	"7/31/2021 13:06"	0.258762	2021-07-31	2021-07-31	1
"11/19/2021 14:06"	"11/19/2021 16:27"	0.130887	2021-11-19	2021-11-19	1
"7/22/2021 10:45"	"7/22/2021 11:31"	0.607945	2021-07-22	2021-07-22	1

```

begin
    dtf = DateFormat("m-d-y")
    Start_date = Vector{Date}{}
    End_date = Vector{Date}{}
    for i in 1:length(df.Start_Time)
        k = findfirst(" ", df.Start_Time[i])
        k2 = findfirst(" ", df.End_Time[i])
        k = k[1]
        k2 = k2[1]
        push!(Start_date, Date(replace((df.Start_Time[i][1:k-1]), "/" => "-"), dtf))
        push!(End_date, Date(replace((df.End_Time[i][1:k2-1]), "/" => "-"), dtf))
    end
    df.Start_Date = Start_date
    df.End_Date = End_date
end

```

### *Julia code snippets 1*

Table 5: Dataset for new road construction projects in US, or projects exceeding 50days of duration

Start_Lng	End_Lat	End_Lng	Distance(mi)	Start_Date	End_Date	Project_Duration_Days	Project_Duration_Class
				Date	Date	Int64	Any
-93.1524	32.8507	-93.1644	1.1035	2019-04-05	2020-09-29	544	"long term"
-83.2652	39.7261	-83.2465	1.10024	2020-04-20	2020-07-22	94	"short term"
-73.9652	40.763	-73.974	0.952071	2021-05-10	2021-12-10	215	"medium term"
-80.1912	26.1871	-80.1874	0.24557	2020-11-15	2021-12-30	411	"long term"
-87.724	41.8312	-87.7242	0.453084	2021-07-04	2021-10-08	97	"short term"
-86.4798	39.6149	-86.4798	0.601812	2021-03-10	2021-12-31	297	"medium term"
-112.063	33.5092	-112.066	0.173348	2021-02-27	2021-12-28	305	"long term"
-80.1906	25.8904	-80.1845	0.380199	2021-04-13	2021-12-13	245	"medium term"
-80.3374	25.7482	-80.3356	0.118257	2021-06-16	2021-12-14	182	"medium term"
-90.7869	42.44	-90.8007	0.701292	2021-05-14	2021-10-22	162	"medium term"
-92.4132	41.0376	-92.4158	0.999269	2020-07-20	2020-10-31	104	"medium term"

## 2.2.1 Feature detection and EDA

Table 6: Extracted features for EDA

	Distance(mi)	Temperature(F)	Wind_Chill(F)	Humidity(%)	Pressure(in)	Visibility(mi)	Wind_Speed(mph)	Precipitation(in)	Project_Duration_Days
1	1.1035	75.0	75.0	58	29.72	10.0	3.0	0.0	544
2	1.10024	70.0	70.0	60	28.89	10.0	5.0	0.0	94
3	0.952071	48.0	46.0	93	29.72	2.0	5.0	0.01	215
4	0.24557	72.0	72.0	78	29.88	10.0	8.0	0.0	411
5	0.453084	64.0	64.0	87	29.37	4.0	0.0	0.0	97
6	0.601812	73.0	73.0	87	29.1	7.0	6.0	0.0	297
7	0.173348	50.0	50.0	31	28.68	10.0	6.0	0.0	305
8	0.380199	71.0	71.0	84	29.95	10.0	0.0	0.0	245
9	0.118257	79.0	79.0	66	30.17	10.0	12.0	0.0	182
10	0.701292	41.0	41.0	65	29.15	10.0	3.0	0.0	162
: more									
43134	0.999269	70.0	70.0	90	29.09	10.0	0.0	0.0	104

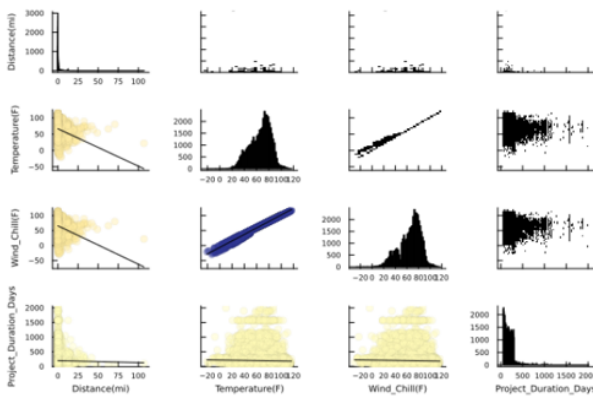


Fig. 1: Correlation plot for Distance, Temperature, Wind chill and Project Duration

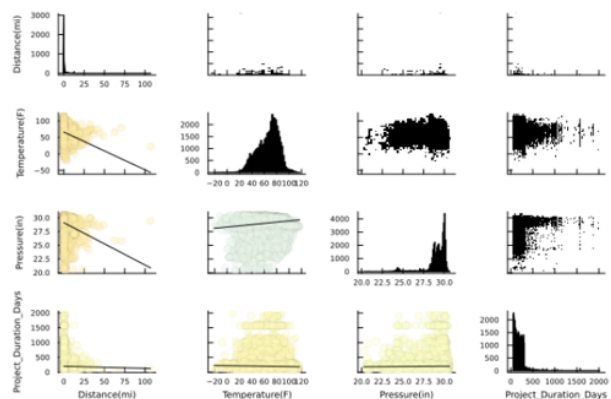


Fig. 2: Correlation plot for Distance, Temperature, Pressure and Project Duration

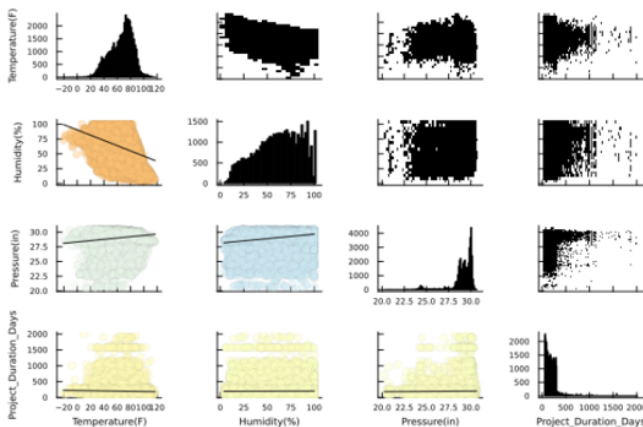


Fig. 3: Correlation plot for Temperature, Humidity, Pressure and Project Duration

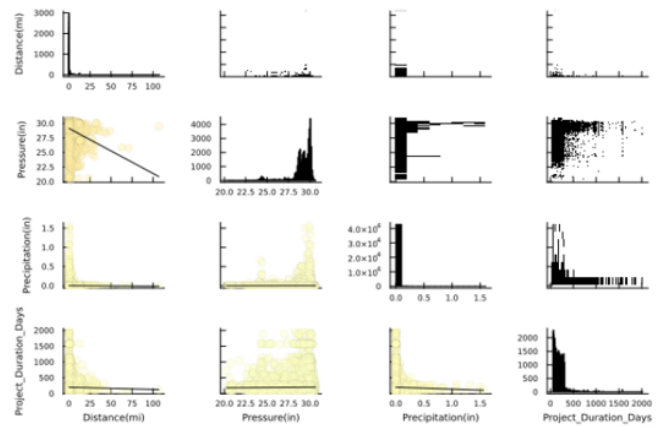


Fig. 4: Correlation plot for Distance, Pressure, Precipitation and Project Duration

```

begin
    arr = collect(1:8)
    plots = []
    combinations_arr = collect(combinations(arr, 3))
    for i in combinations_arr
        push!(i, 9)
        push!(plots, @df df_sel corrplot(cols(i), grid = false, size(10000,8000),
            xtickfontsize=5, ytickfontsize=5, xguidefontsize=6, yguidefontsize=6))
    end
    plots
end
end

```

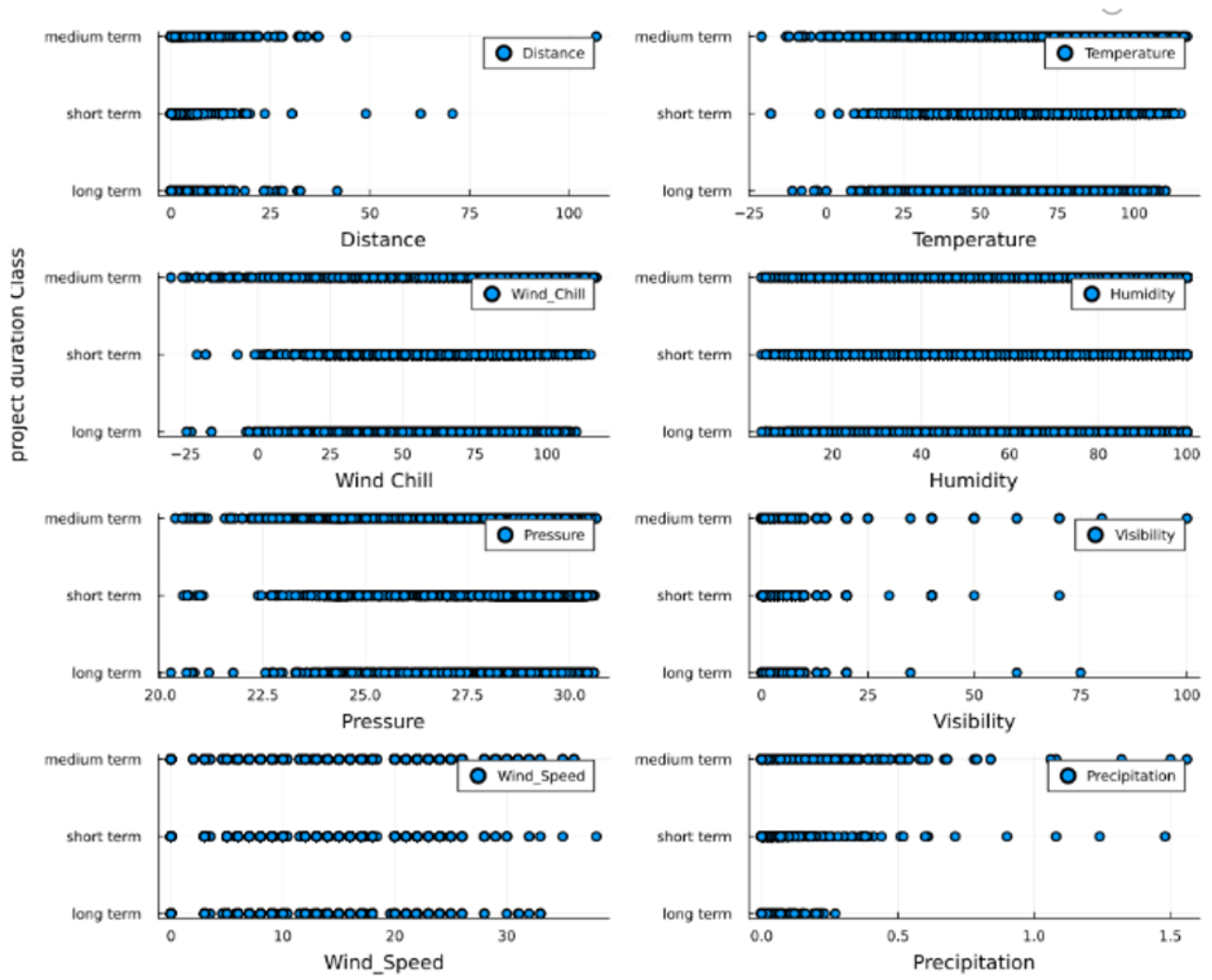
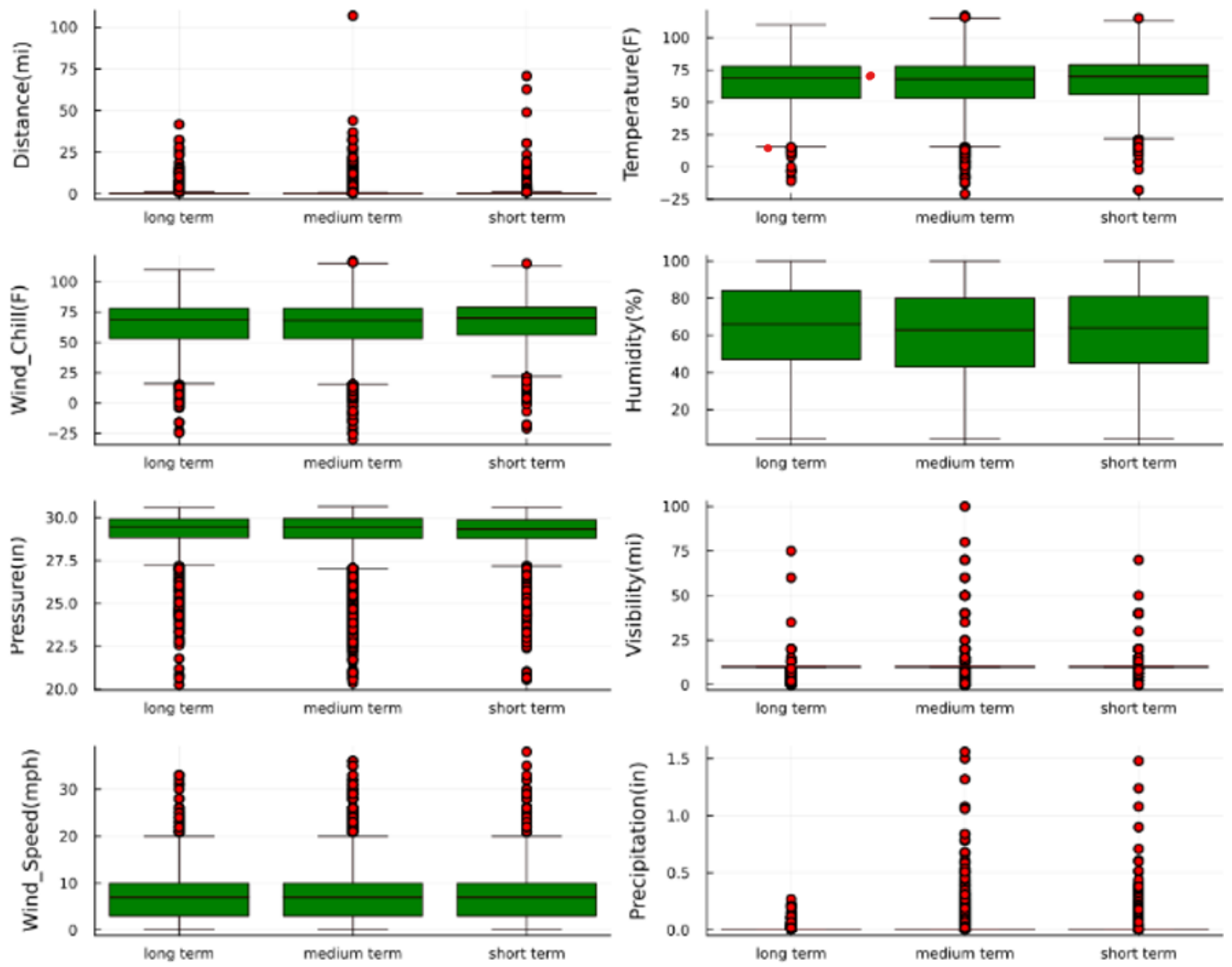


Fig. 5: Scatter plots of pertinent features against Project Duration Class



*Fig. 6: Box plots for outlier detection in pertinent features*

## 2.3 Dimensionality reduction

	Number_of_Modes	Fraction_of_Variance
<b>1</b>	1	0.640433
<b>2</b>	2	0.970394
<b>3</b>	3	0.991476
<b>4</b>	4	0.995204
<b>5</b>	5	0.9978
<b>6</b>	6	0.998981
<b>7</b>	7	0.999999
<b>8</b>	8	1.0

```
begin
    var_pca = [i^2 for i in F.S]
    frac_i = []
    n_modes = []
    for i in 1:length(var_pca)
        push!(frac_i, sum(var_pca[1:i])/sum(var_pca))
        push!(n_modes, i)
    end

    PCA = DataFrame(Number_of_Modes = n_modes, Fraction_of_Variance=frac_i)
end
```

*Julia code snippets 3*



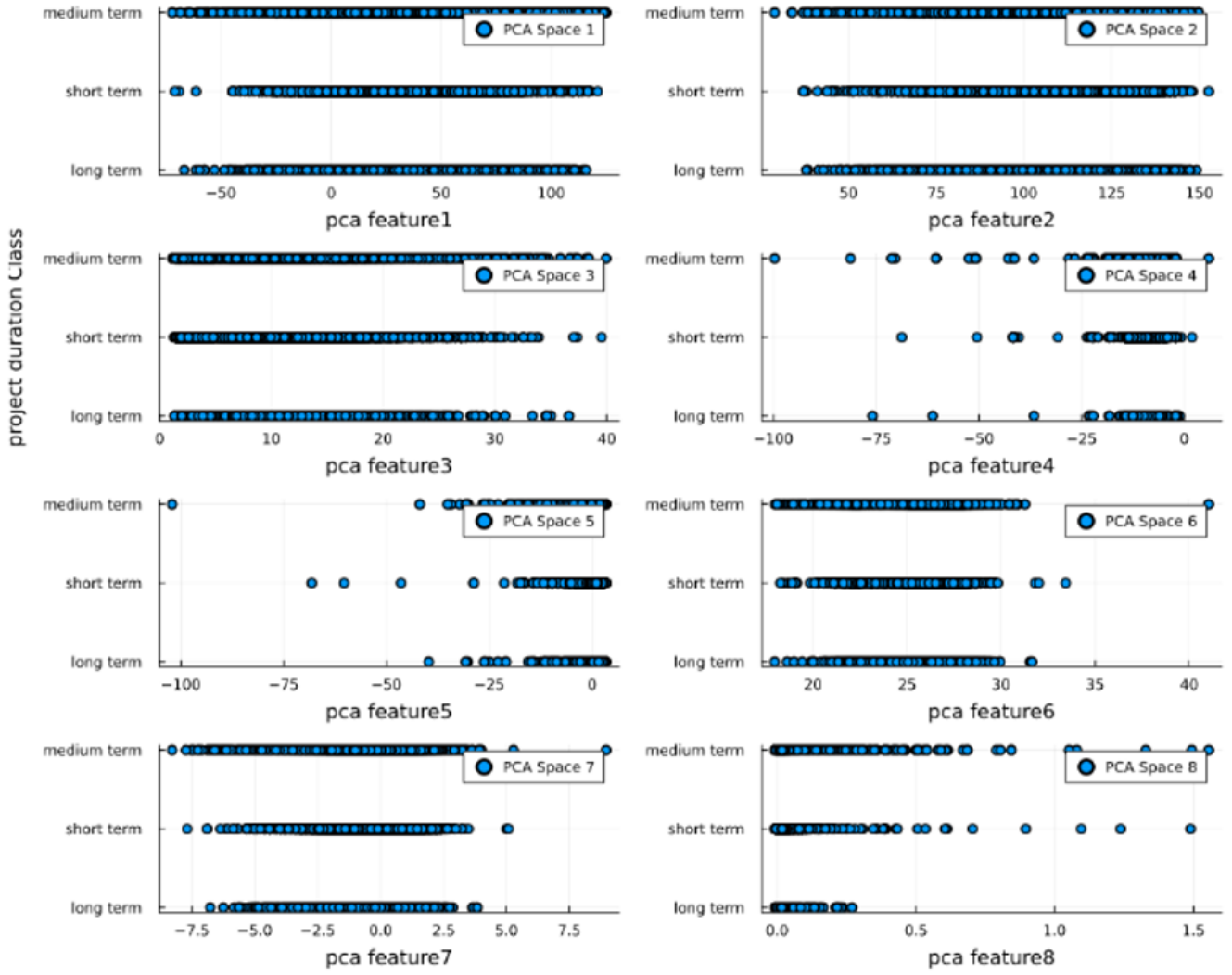
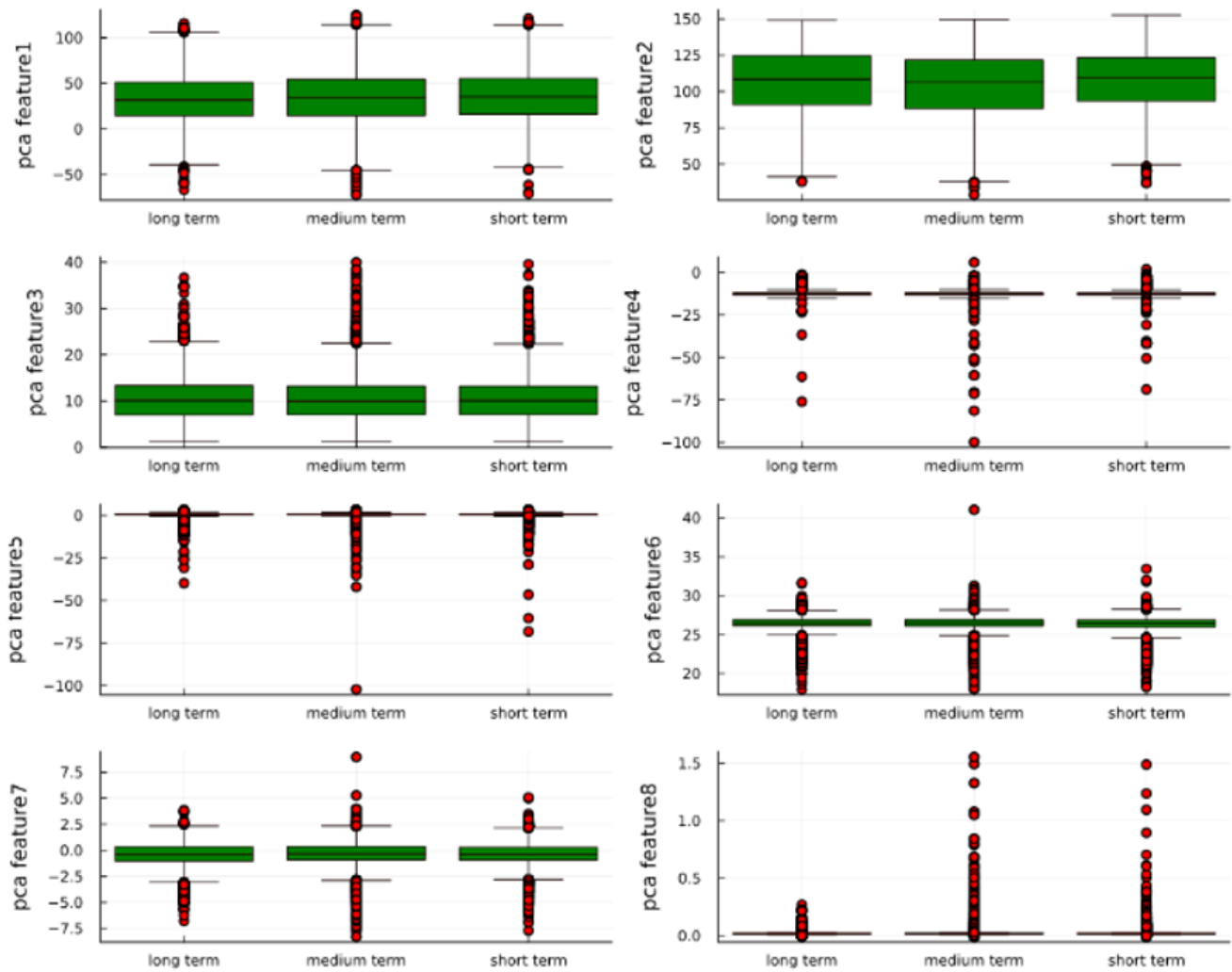


Fig. 7: Transformed data in PCA coordinate frame



*Fig. 8: Boxplots of transformed data in PCA coordinate frame*

### 3.0. Predictive model development

#### References:

1. Karimi Monsefi, Amin, Sobhan Moosavi, and Rajiv Ramnath. "Will there be a construction? Predicting road constructions based on heterogeneous spatiotemporal data.", 2022
2. US Road Construction and Closures 2016 - 2021 from Kaggle