DataLimn

GET INSIGHT IN YOUR DATA

- **Home**
- UserLogin
- UploadData
- VisualizeData
- MachineLearning

Welcome to DataLimn!

DataLimn is your go-to open-source framework for machine learning and data science projects. The easy to use app gives greater insight and detailed information about your data. You simply upload a CSV file, follow some steps, and get your results!

What does DataLimn do?

DataLimn is a tool designed to analyse patterns and make predictions without having to have previous knowledge of coding. This no-code machine learning app evaluates and interprets your data, transforming it into an understandable story that can be of great value in business decision-making processes.

Module description

Data uploading and preparation

Upload a CSV file with your data. Once the data is uploaded the user can change the data type per column and select the training size. Furthermore , select the columns you would want to take into consideration as well as the column where you want to base your predictions on.

Machine learning and data visualisation

The next step is to select one (or multiple) learning model(s) based on your data and the outcome you would like to see. Now it is time for our app to do the hard work. The app will give you predictions and probabilities based on your data input, and some statistical graphs can be selected. Moreover, you can modify your input and generate a new outcome based on the features given.

next

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

💬 User Login

| ▷ **Register** | ▷ Login | ▷ Update password |

Expand to register ⌃

Enter your name

Enter your email address

Enter a password 👁

Register

next

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Upload Your File

Please enter tags of your data separating by comma

Choose CSV file(s)

☁️ **Drag and drop files here**
Limit 200MB per file • CSV

Browse files

next

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Upload Your File

Please enter tags of your data separating by comma

diabetes, pregnancies

Choose CSV file(s)

Drag and drop files here
Limit 200MB per file • CSV

Browse files

Diabetes_dataset.csv 23.9KB

Details of Your File

```
{
    "_id" : "a008c79d519e42a88966dbc622757ae8"
    "File Name" : "Diabetes_dataset.csv"
    "File Size" : 23873
    "File Tags" : "diabetes, pregnancies"
}
```

Your Dataframe

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	6	148	72	35	0	33.6000	0.6270	5(
1	1	85	66	29	0	26.6000	0.3510	3:
2	8	183	64	0	0	23.3000	0.6720	3:
3	1	89	66	23	94	28.1000	0.1670	2:
4	0	137	40	35	168	43.1000	2.2880	3:

4

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

General Information of Your Data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

Description of Your Data

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunctio
count	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000	768.000
mean	3.8451	120.8945	69.1055	20.5365	79.7995	31.9926	0.471
std	3.3696	31.9726	19.3558	15.9522	115.2440	7.8842	0.331
min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.078
25%	1.0000	99.0000	62.0000	0.0000	0.0000	27.3000	0.243
50%	3.0000	117.0000	72.0000	23.0000	30.5000	32.0000	0.372
75%	6.0000	140.2500	80.0000	32.0000	127.2500	36.6000	0.626
max	17.0000	199.0000	122.0000	99.0000	846.0000	67.1000	2.420

Dropping Duplicates

Number of rows is : 768

After dropping duplicates, number of rows is: 768

Splitting Strings and Integers

Your data may contain some columns which have both string and numbers. For example: 10 km

I must remove strings in these columns to run numerical calculations. Please help me !

Choose all columns name to fix them:

Select multiple columns to fix

Choose an option	▼

Please note that you see default dataframe in the beginning. This dataframe will be restored as soon as your selection is done.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	↑ DiabetesPedigreeFunction	Ag
268	0	102	52	0	0	25.1000	0.0780	2
180	6	87	80	0	0	23.2000	0.0840	3
149	2	90	70	17	0	27.3000	0.0850	2
567	6	92	62	32	126	32.0000	0.0850	4
135	2	125	60	20	140	33.8000	0.0880	3
598	1	173	74	0	0	36.8000	0.0880	3
229	0	117	80	31	53	45.2000	0.0890	2
257	2	114	68	22	0	28.7000	0.0920	2
146	9	57	80	37	0	32.8000	0.0960	4
602	1	124	74	36	0	27.8000	0.1000	3

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Handling Missing Values

All missing values are restored by median value of related column.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Ag
0	6	148	72	35	0	33.6000	0.6270	5
1	1	85	66	29	0	26.6000	0.3510	3
2	8	183	64	0	0	23.3000	0.6720	3
3	1	89	66	23	94	28.1000	0.1670	2
4	0	137	40	35	168	43.1000	2.2880	3
5	5	116	74	0	0	25.6000	0.2010	3
6	3	78	50	32	88	31.0000	0.2480	2
7	10	115	0	0	0	35.3000	0.1340	2
8	2	197	70	45	543	30.5000	0.1580	5
9	8	125	96	0	0	0.0000	0.2320	5

General Information of Your Data After Missing Values Handling

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   Pregnancies               768 non-null     int64
 1   Glucose                   768 non-null     int64
 2   BloodPressure             768 non-null     int64
 3   SkinThickness             768 non-null     int64
 4   Insulin                   768 non-null     int64
 5   BMI                       768 non-null     float64
 6   DiabetesPedigreeFunction  768 non-null     float64
```

General Information of Your Data After Missing Values Handling

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 60.0 KB
```

Download Your Cleaned Data

Data handling process is done! Please save the new clean data for the next steps!!

Press to Download

next

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Get Plots of Your Data

Please upload your cleaned data which you downloaded in the "Upload Data" section.

Choose CSV file(s)

☁️ **Drag and drop files here**
Limit 200MB per file • CSV

Browse files

next

Get Plots of Your Data

Please upload your cleaned data which you downloaded in the "Upload Data" section.

Choose CSV file(s)

☁ **Drag and drop files here**
Limit 200MB per file • CSV

Browse files

📄 Diabetes_dataset.csv 23.9KB ✕

Please select a single variable for Pie Plot

Select a value

Pregnancies	▼

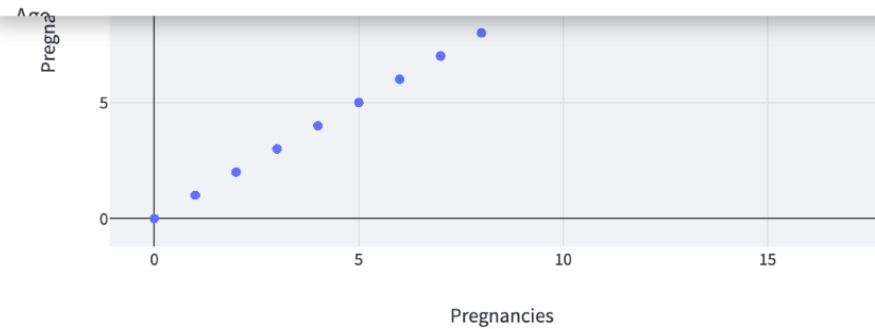| Pregnancies |
| --- |
| Glucose |
| BloodPressure |
| SkinThickness |
| Insulin |
| BMI |
| DiabetesPedigreeFunction |
| Age |

DataLimn
GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData
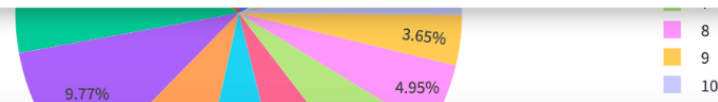
MachineLearning

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Choose CSV file(s)

Drag and drop files here
Limit 200MB per file • CSV

Browse files

Diabetes_dataset.csv 23.9KB

Please select a single variable for Pie Plot

Select a value

Pregnancies

Pie Plot
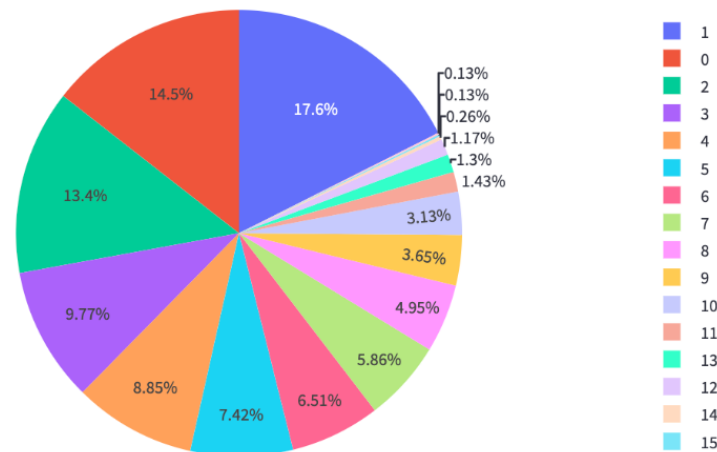
DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

## Please select x and y variables for Scatter Plot, Line Plot, Bar Plot and Box Plot

Select x values

| Pregnancies | ▼ |

Select y value

| Pregnancies| | ▼ |

| Pregnancies |
| Glucose |
| BloodPressure |
| SkinThickness |
| Insulin |
| BMI |
| DiabetesPedigreeFunction |
| Age |

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Please select x and y variables for Scatter Plot, Line Plot, Bar Plot and Box Plot

Select x values

Pregnancies ▼
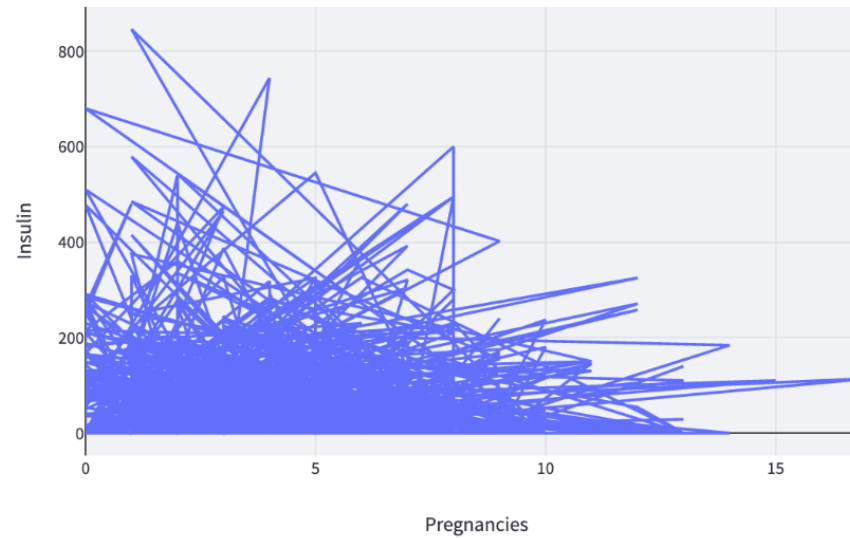
Select y value

Insulin ▼

Scatter Plot
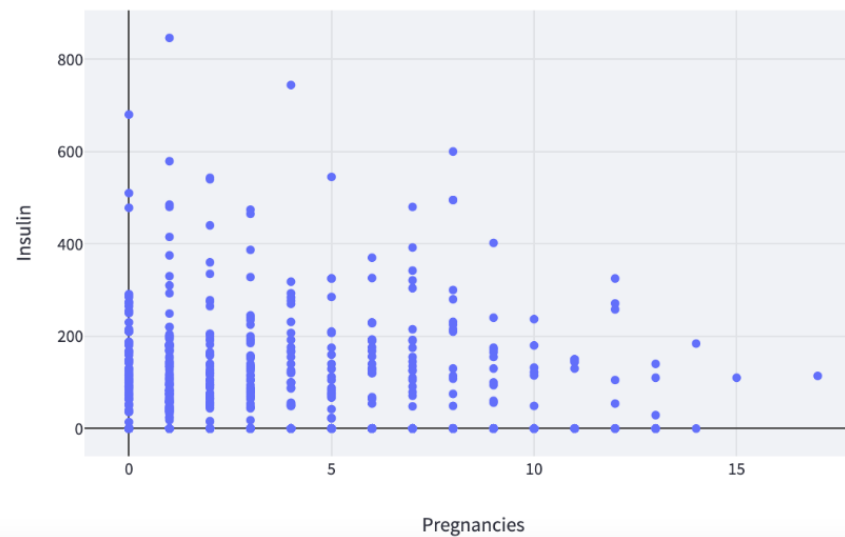
Home

UserLogin

UploadData
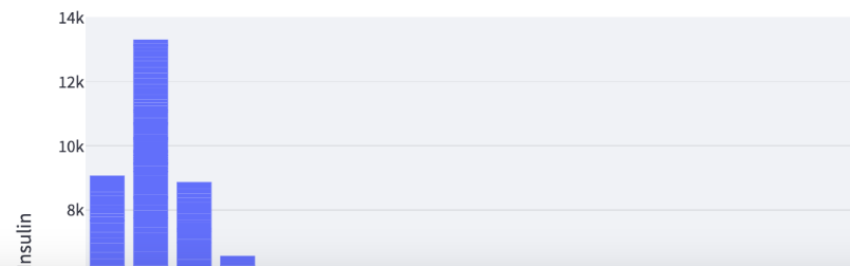
VisualizeData

MachineLearning

Line Plot

View fullscreen

Bar Plot

Box Plot

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Pregnancies

Box Plot

(6, max: 370)

(6, upper fence: 230)

(6, q3: 130)

(6, median: 0)

(6, q1: 0)

(6, lower fence: 0)

(6, min: 0)

Pregnancies
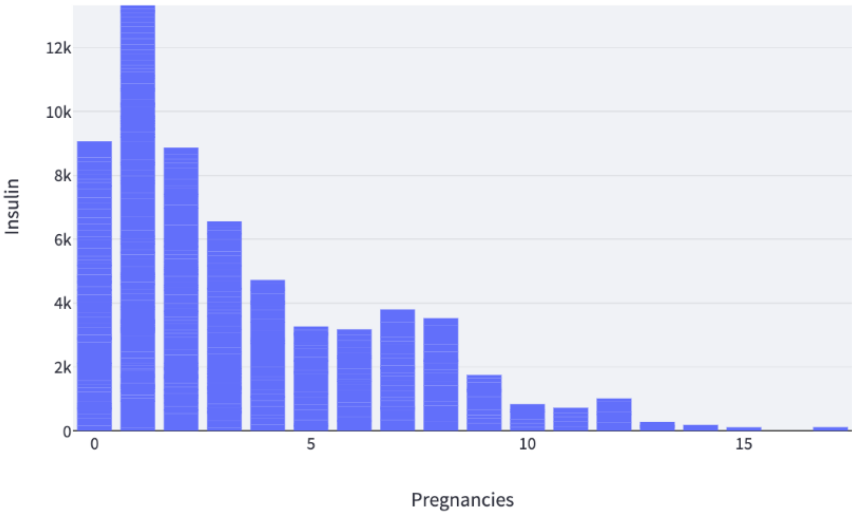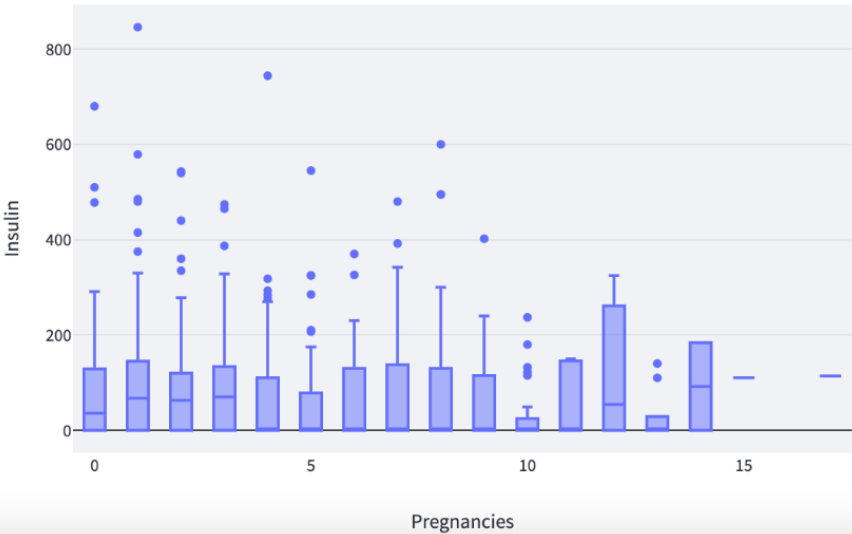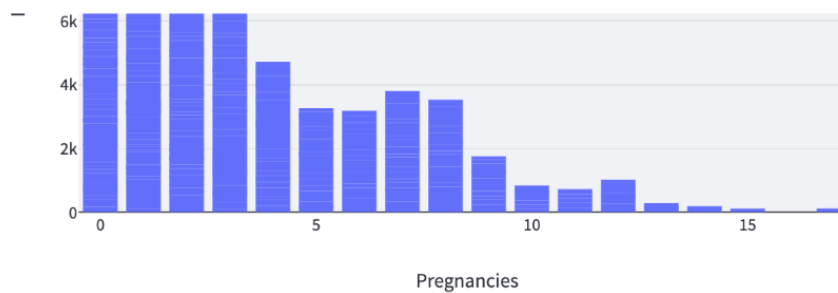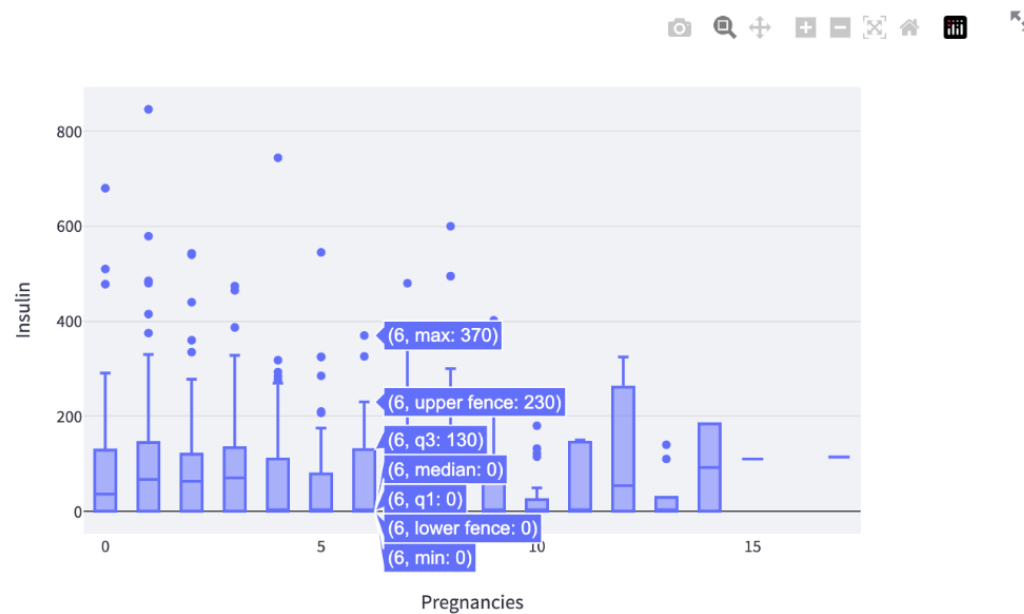
next

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Machine Learning

Please upload your cleaned data which you downloaded in the "Upload Data" section.

Choose CSV file(s)

Drag and drop files here   Browse files
Limit 200MB per file • CSV

Select the variable to be predicted (y)

Pregnancies   ▼

Select the variables to be used for prediction (X)

Choose an option   ▼

You have to put in some X variable and it cannot be left empty.

Select the type of process you want to run. ⓘ

◉ Linear regression
◯ Logistic regression
◯ KNN
◯ Gaussian Naive Bayes
◯ SVM
◯ Decision tree
◯ Random forest

Run Models

DataLimn

GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

×

≡

Please upload your cleaned data which you downloaded in the "Upload Data" section.

Choose CSV file(s)

☁ **Drag and drop files here**
Limit 200MB per file • CSV

Browse files

Select the variable to be predicted (y)

Pregnancies ▾

Select the variables to be used for prediction (X)

Glucose × | SkinThickness × | Insulin × | BMI × | Age × | ⊗ ▾

Pregnancies

BloodPressure

DiabetesPedigreeFunction

Outcome

○ SVM

○ Decision tree

○ Random forest

Run Models

Train Test Splitting

Select training size ⊘

0.74

0.00 ————————————————————————————————————— 1.00

Number of training samples: 568

Number of testing samples: 200

18

DataLimn
GET INSIGHT IN YOUR DATA

Home

UserLogin

UploadData

VisualizeData

MachineLearning

Select the type of process you want to run. (?)

○ Linear regression
● Logistic regression
○ KNN
○ Gaussian Naive Bayes
○ SVM
○ Decision tree
○ Random forest

[Run Models]

Train Test Splitting

Select training size (?)

0.74

0.00   1.00

Number of training samples: 568

Number of testing samples: 200

Logistic Regression Results

Confusion Matrix:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	13	6	0	0	2	1	0	0	0	1	0	0	0	0
1	15	15	0	1	0	1	1	1	0	0	0	0	1	0
2	18	10	0	0	0	0	0	0	1	0	0	0	0	0
3	10	4	0	0	1	1	1	2	1	0	0	0	0	0
4	10	10	0	0	0	2	1	0	1	0	0	0	0	0
5	1	6	0	0	2	1	2	1	0	0	1	0	0	0
6	1	2	0	0	1	2	1	1	0	0	0	0	0	0

19

5	1	6	0	0	2	1	2	1	0	0	1	0	0	0
6	1	2	0	0	1	2	1	1	0	0	0	0	0	0
7	1	3	0	0	3	3	1	1	1	0	0	0	0	0
8	0	3	0	0	3	6	0	0	0	0	0	0	0	0
9	0	3	0	0	1	3	1	0	0	0	0	0	0	0

Classification Report: precision recall f1-score support

```
0       0.18       0.57       0.27       23
1       0.23       0.43       0.30       35
2       0.00       0.00       0.00       29
3       0.00       0.00       0.00       20
4       0.00       0.00       0.00       24
5       0.05       0.07       0.06       14
6       0.11       0.12       0.12        8
7       0.12       0.08       0.10       13
8       0.00       0.00       0.00       12
9       0.00       0.00       0.00        8
10      0.00       0.00       0.00        5
11      0.00       0.00       0.00        3
12      0.00       0.00       0.00        3
13      0.00       0.00       0.00        3

accuracy                      0.15      200
```

macro avg 0.05 0.09 0.06 200 weighted avg 0.08 0.15 0.10 200

Accuracy Score: `0.155`

20