ANALYTICS FOR HOSPITALS HEALTH CARE DATA

TEAM ID :PNT2022TMID37727

PREDICTION OF LENGTH OF STAY

SPRINT 4

```
!pip install pyspark
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Collecting pyspark
  Downloading pyspark-3.3.1.tar.gz (281.4 MB)
    |████████████████████████████████| 281.4 MB 47 kB/s
Collecting py4j==0.10.9.5
  Downloading py4j-0.10.9.5-py2.py3-none-any.whl (199 kB)
    |████████████████████████████████| 199 kB 35.4 MB/s
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.3.1-py2.py3-none-any.whl size=281845514
  Stored in directory: /root/.cache/pip/wheels/42/59/f5/79a5bf931714dcd201b26025347785f
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9.5 pyspark-3.3.1
```

```
from google.colab import files
uploaded = files.upload()
!pip install pyspark
from pyspark.sql import SparkSession
import seaborn as sns
import matplotlib.pyplot as plt

spark = SparkSession.builder.master('local')\
.appName("Predicting LOS for High Risk Patient")\
.getOrCreate()
```

Choose Files train_data.csv

- **train_data.csv**(text/csv) - 26915586 bytes, last modified: 8/23/2021 - 100% done

```
Saving train_data.csv to train_data (2).csv
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Requirement already satisfied: pyspark in /usr/local/lib/python3.7/dist-packages (3.3.1
Requirement already satisfied: py4j==0.10.9.5 in /usr/local/lib/python3.7/dist-packages
```

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
import pandas as pd
import io
```

```
df = pd.read_csv(io.BytesIO(uploaded['train_data.csv']))
print(df)
```

```
318434   318435           24                  a                   1
318435   318436            7                  a                   4
318436   318437           11                  b                   2
318437   318438           19                  a                   7

       Hospital_region_code  Available Extra Rooms in Hospital    Department  \
0                         Z                                  3  radiotherapy   
1                         Z                                  2  radiotherapy   
2                         X                                  2    anesthesia   
3                         Y                                  2  radiotherapy   
4                         Y                                  2  radiotherapy   
...                     ...                                ...           ...   
318433                    X                                  3  radiotherapy   
318434                    X                                  2    anesthesia   
318435                    X                                  3    gynecology   
318436                    Y                                  3    anesthesia   
318437                    Y                                  5    gynecology   

       Ward_Type Ward_Facility_Code  Bed Grade  patientid  City_Code_Patient  \
0              R                  F        2.0      31397                7.0   
1              S                  F        2.0      31397                7.0   
2              S                  E        2.0      31397                7.0   
3              R                  D        2.0      31397                7.0   
4              S                  D        2.0      31397                7.0   
...          ...                ...        ...        ...                ...   
318433         Q                  F        4.0      86499               23.0   
318434         Q                  E        4.0        325                8.0   
318435         R                  F        4.0     125235               10.0   
318436         Q                  D        3.0      91081                8.0   
318437         Q                  C        2.0      21641                8.0   

       Type of Admission Severity of Illness  Visitors with Patient    Age  \
0              Emergency             Extreme                      2  51-60   
1                 Trauma             Extreme                      2  51-60   
2                 Trauma             Extreme                      2  51-60   
3                 Trauma             Extreme                      2  51-60   
4                 Trauma             Extreme                      2  51-60   
...                  ...                 ...                    ...    ...   
318433         Emergency            Moderate                      3  41-50   
318434            Urgent            Moderate                      4  81-90   
318435         Emergency               Minor                      3  71-80   
318436            Trauma               Minor                      5  11-20   
318437         Emergency               Minor                      2  11-20   

        Admission_Deposit   Stay  
0                  4911.0   0-10  
1                  5954.0  41-50  
2                  4745.0  31-40  
3                  7272.0  41-50  
4                  5558.0  41-50  
...                   ...    ...  
318433             4144.0  11-20  
318434             6699.0  31-40  
```

```
     318434              6699.0  31-40
     318435              4235.0  11-20
     318436              3761.0  11-20
     318437              4752.0   0-10

     [318438 rows x 18 columns]
```

```
!pip install -q findspark
!pip install pyspark
!pip install matplotlib
!pip install seaborn
```

```
     Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publ
     Requirement already satisfied: pyspark in /usr/local/lib/python3.7/dist-packages (3.3.1)
     Requirement already satisfied: py4j==0.10.9.5 in /usr/local/lib/python3.7/dist-packages
     Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publ
     Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (3.2
     Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pac
     Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/li
     Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.7/dist-packages (fr
     Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packag
     Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (f
     Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packag
     Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from
     Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publ
     Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (0.11.2
     Requirement already satisfied: pandas>=0.23 in /usr/local/lib/python3.7/dist-packages (f
     Requirement already satisfied: matplotlib>=2.2 in /usr/local/lib/python3.7/dist-packages
     Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.7/dist-packages (fr
     Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.7/dist-packages (fro
     Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pac
     Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packag
     Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (f
     Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/li
     Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packag
     Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (f
     Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from
```

```
import findspark
findspark.find()
```

```
     '/usr/local/lib/python3.7/dist-packages/pyspark'
```

```
from pyspark.sql import SparkSession
import seaborn as sns
import matplotlib.pyplot as plt

spark = SparkSession.builder.master('local')\
.appName("Predicting LOS for High Risk Patient")\
.getOrCreate()
```

```
spark
```

SparkSession - in-memory

SparkContext

Spark UI

Version
```
v3.3.1
```
Master
```
local
```
AppName
```
Predicting LOS for High Risk Patient
```

```
print(f"Counts of rows/samples: {df.count()}")
print(f"Counts of columns/features: {len(df.columns)}")
```

```
Counts of rows/samples: case_id                              318438
Hospital_code                        318438
Hospital_type_code                   318438
City_Code_Hospital                   318438
Hospital_region_code                 318438
Available Extra Rooms in Hospital    318438
Department                           318438
Ward_Type                            318438
Ward_Facility_Code                   318438
Bed Grade                            318325
patientid                            318438
City_Code_Patient                    313906
Type of Admission                    318438
Severity of Illness                  318438
Visitors with Patient                318438
Age                                  318438
Admission_Deposit                    318438
Stay                                 318438
dtype: int64
Counts of columns/features: 18
```

```
df
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region
0	1	8	c	3	
1	2	2	c	5	
2	3	10	e	1	
3	4	26	b	2	
4	5	26	b	2	
...	...	...	...	...	
318433	318434	6	a	6	
318434	318435	24	a	1	
318435	318436	7	a	4	

```
input_variable = ['hospital', 'hospital_type', 'hospital_city','hospital_region','available_e
                  'bed_grade','city_code_patient','patient_visitors','admission_deposit',
                  'department_index', 'ward_facility_index', 'ward_type_index', 'illness_seve
                  'type_of_admission_index']

label = ['stay_days_index']
```

```
from pyspark.ml.feature import PCA

pca =PCA(k=10, inputCol="features", outputCol="pcaFeatures")
```

```
from pyspark.ml.feature import StandardScaler

scaler = StandardScaler(inputCol="pcaFeatures", outputCol="scaledFeatures",
                        withStd=True, withMean=False)
```

```
# pipeline = Pipeline(stages=[])
```

```
df.corr().style.background_gradient(cmap='coolwarm').set_precision(2)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: this meth
  """Entry point for launching an IPython kernel.
```

	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patien
case_id	1.00	-0.04	-0.01	0.04	0.01	-(
Hospital_code	-0.04	1.00	0.13	-0.06	-0.01	(
City_Code_Hospital	-0.01	0.13	1.00	-0.05	-0.05	(
Available Extra Rooms in Hospital	0.04	-0.06	-0.05	1.00	-0.12	(
Bed Grade	0.01	-0.01	-0.05	-0.12	1.00	(
patientid	-0.00	0.00	0.00	0.00	0.00	'
City_Code_Patient	0.07	-0.02	-0.02	-0.01	-0.01	(
Visitors with						

```
df.corr().style.background_gradient(cmap='coolwarm').set_precision(2)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: this meth
  """Entry point for launching an IPython kernel.
```

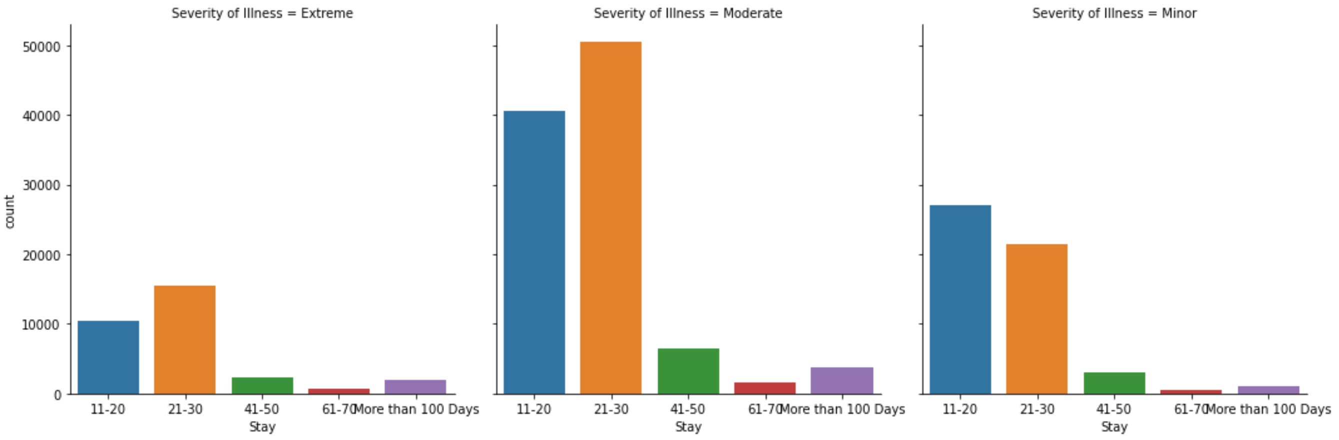
	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patien
case_id	1.00	-0.04	-0.01	0.04	0.01	-(
Hospital_code	-0.04	1.00	0.13	-0.06	-0.01	(
City_Code_Hospital	-0.01	0.13	1.00	-0.05	-0.05	(
Available Extra Rooms in Hospital	0.04	-0.06	-0.05	1.00	-0.12	(
Bed Grade	0.01	-0.01	-0.05	-0.12	1.00	(
patientid	-0.00	0.00	0.00	0.00	0.00	'
City_Code_Patient	0.07	-0.02	-0.02	-0.01	-0.01	(
Visitors with						

```
selected_list = ["11-20","21-30","41-50","61-70", "More than 100 Days"]

def bivariate_analysis(dataframe, dependent_variable, independent_variable, selected_list):
   g = sns.catplot(dependent_variable, col=independent_variable, col_wrap=3,\
   data=dataframe,kind="count", height=5, aspect=1, order=selected_list
   )
```

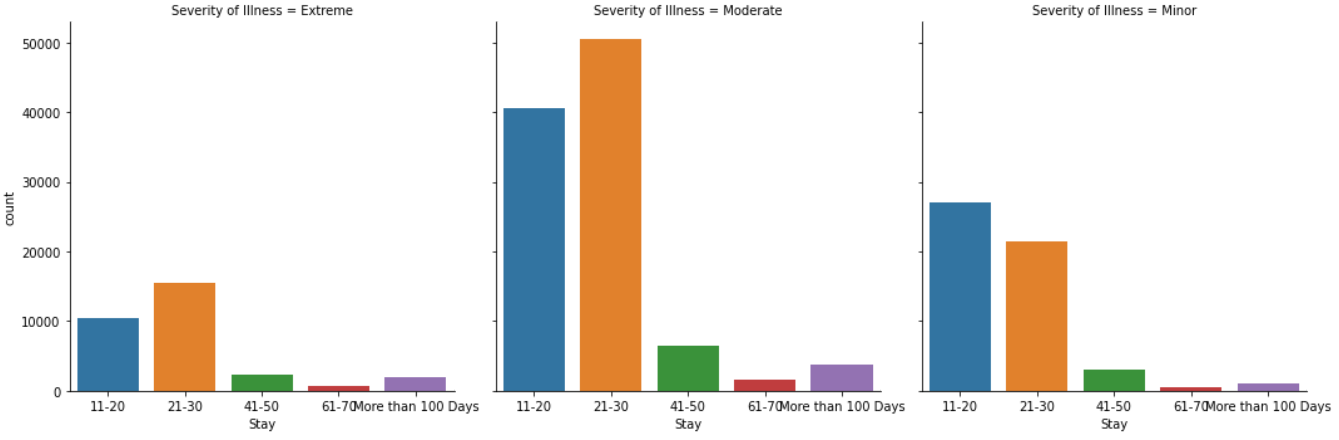
```
bivariate_analysis(df, "Stay", "Severity of Illness", selected_list)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass th
  FutureWarning
```



```
bivariate_analysis(df, "Stay", "Severity of Illness", selected_list)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass th
  FutureWarning
```



✓ 0s completed at 11:08 PM ● ×