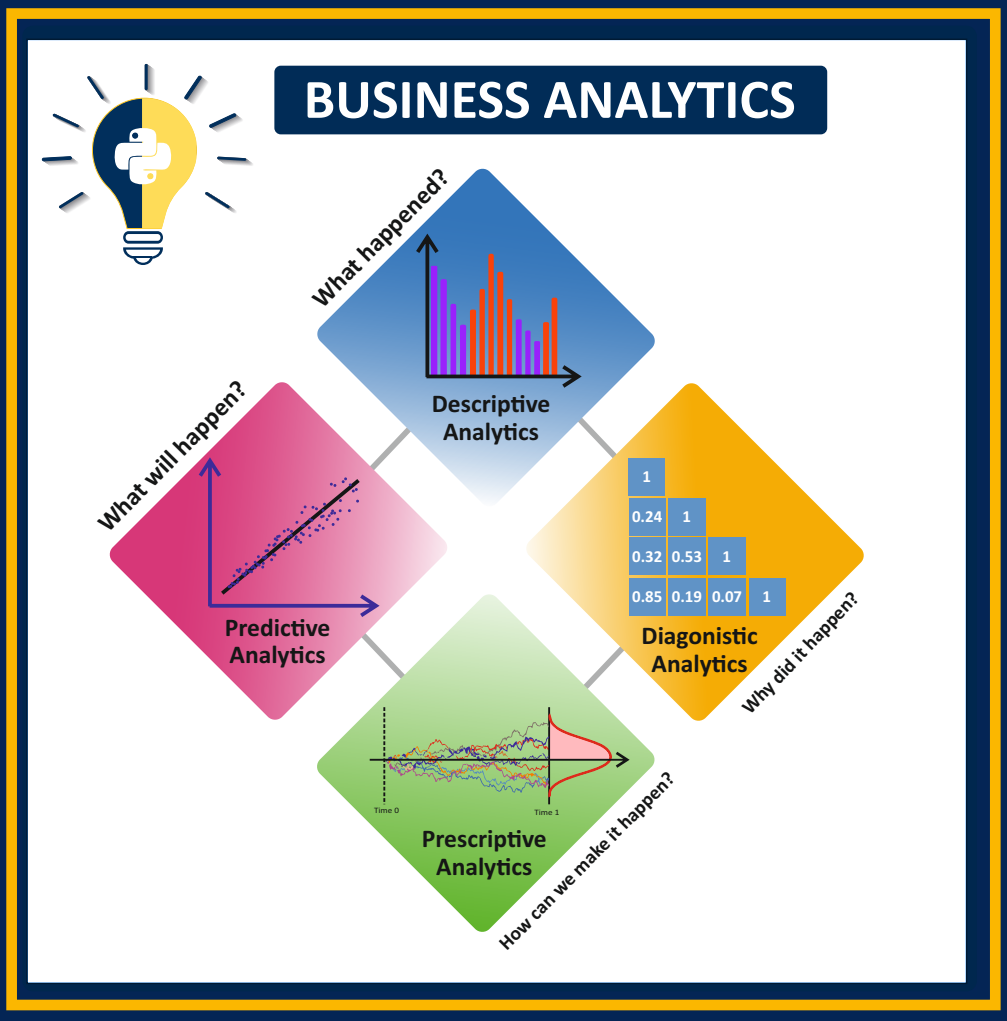
BUSINESS ANALYTICS
using Python

50+ hours

Case Study and Project- driven Methodology

Blended Learning Methodology



PEAKS2TAILS

DETAILED CURRICULUM

MODULE 1 -FEATURE ENGINEERING

STEPS OF DATA CLEANING & PROCESSING

- Identifying Data type
- Exploratory data analysis - Data Summarisation and visualisation
- Missing value treatment for Categorical Variables
- Missing value treatment for Continuous Variables
- Outliers treatment for Categorical Variables
- Outliers treatment for Countinuous Variables
- Balancing of data
- Covariaties Creation (Enrichment)
- Dimensions Reductions
- Dummy Coding for Categorical Variables
- Scaling for continuous Variables
- Discretisation or Weight of Evidence
- Data Partitioning

MODULE 2 -MODEL BUILDING

STEPWISE REGRESSION

- Removing problems of Multicollinearity,
- Selecting Important Variables
- Correcting Problems of autocorrelation
- Checking problems of Non-Linearity
- Then correcting the Problems of Heteroskedasticity

DETAILED CURRICULUM

TYPES OF REGRESSION	▪ Multiple Regression ▪ Dummy Independent Variable - Dummy Regression ▪ Dummy Dependant Variable - Logistic Regression ▪ Penalised Regression - Ridge & Lasso Regression ▪ Forward Selection, Forward stagewise and least angle Regression ▪ Bayesian regression with Spike & Slab Selection ▪ Support Vector Regression ▪ Principal Component Regression
LOGISTIC REGRESSION	▪ General Linear Modelling in Excel using 4 main link functions Normal, Poisson, Binomial, Gamma. ▪ Multinomial Logit, Ordered Logit Model
DECISION TREE	▪ Classification and Decision Tree ▪ Classification & Regression Tree ▪ CHAID ▪ Boosting, Bagging & Random Forest
SUPPORT VECTOR MACHINE	▪ Primal & Dual Formulation ▪ Linear SVM, Non Linear SVM using Slack Variables ▪ Kernel Trick and Radial Basis Function
LINEAR DISCRIMINANT ANALYSIS	▪ Maximum Likelihood ▪ Fisher's Discriminant ▪ Bayesian Discrimination
K -NEAREST NEIGHBOUR	▪ 1.K-Nearest neighbour ▪ 2.K-means protype

DETAILED CURRICULUM

NAIVE BAYES CLASSIFIER	▪ Bayes Theorem ▪ Naïve Bayes Classifier
CLUSTERING	▪ Bottom up Clustering a.k.a K means clustering ▪ Bottom up Clustering a.k.a Hierarchical clustering ▪ Top down Clustering a.k.a Minimal Spanning Tree ▪ Clustering using Expectation Maximization ▪ Mixed Variables Clustering - K means Prototype

MODULE 3 - MODELLING VALIDATION

VALIDATION METRICS	▪ Gini/AR ▪ AUROC/ CAP ▪ Unconditional Entropy ▪ Conditional Entropy ▪ Kullback-Leibler Divergence ▪ Kolmogorov-Smirnov (KS) ▪ Information Value

MODULE 4 - PROJECTS

PROJECTS	▪ Application to Credit data ▪ Application to HR data ▪ Application to Fraud detection ▪ Application to Marketing data

BACKGROUND | OBJECTIVE | ATTENDEES | PEDADOGY

BACKGROUND

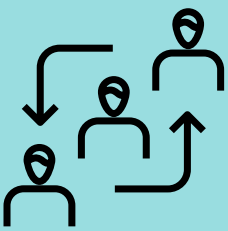
The amount of data that is being generated on a daily basis has increased multiple times, the technological advancement in terms of storing this data has improved considerably. Therefore, in turn the ability to analyse the data and generate insights for data driven decision making becomes primarily importance However the availability of manpower with data science and machine learning skills is limited. The objective of the course is to introduce the concepts of data science and machine learning to the participants using python.

OBJECTIVE

Develop predictive models using various statistical and machine learning techniques, Interpret and evaluate various models and its generalization, Hands on experience on the usage of open notebooks in Python like Jupiter.

WHO CAN ATTEND

Beginner candidates from various quantitative backgrounds, like Engineering, Finance, Maths, Business Management who are looking for Business Analytics training to start their career in the field of Analytics and Data Science.

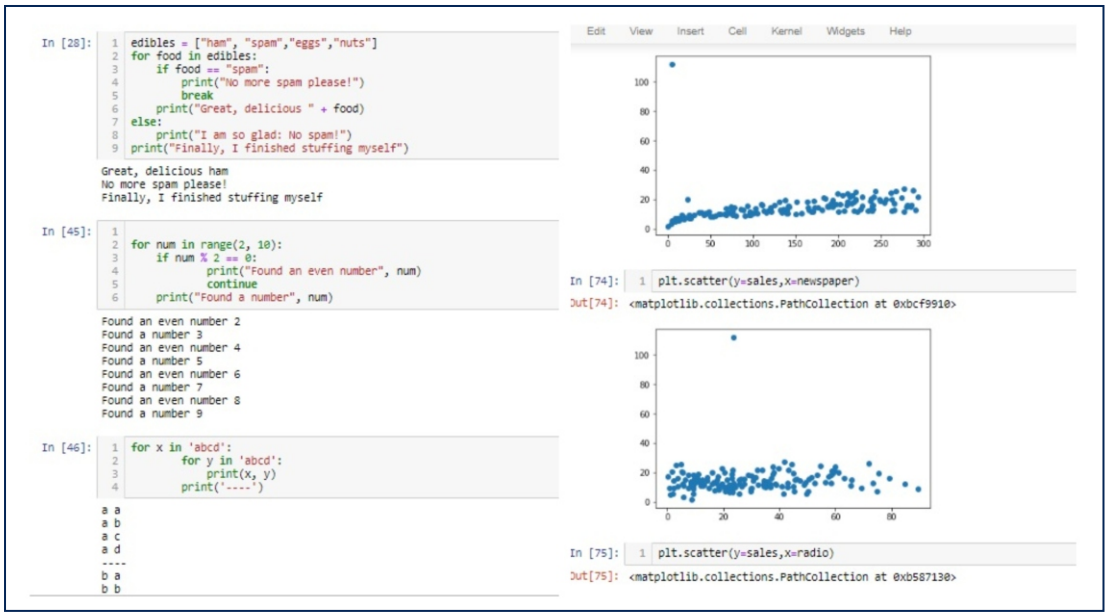


PEDADOGY

Professionals working in analytics field or students interested to make a career in analytics

DEMO MODELS

DURING THE PROGRAM YOU WILL LEARN TO CREATE EXCEL MODELS LIKE SHOWN BELOW



```
In [28]:
edibles = ["ham", "spam","eggs","nuts"]
for food in edibles:
    if food == "spam":
        print("No more spam please!")
        break
    print("Great, delicious " + food)
else:
    print("I am so glad: No spam!")
print("Finally, I finished stuffing myself")
```

```
Great, delicious ham
No more spam please!
Finally, I finished stuffing myself
```

```
In [45]:
for num in range(2, 10):
    if num % 2 == 0:
            print("Found an even number", num)
            continue
    print("Found a number", num)
```

```
Found an even number 2
Found a number 3
Found an even number 4
Found a number 5
Found an even number 6
Found a number 7
Found an even number 8
Found a number 9
```

```
In [46]:
for x in 'abcd':
        for y in 'abcd':
            print(x, y)
        print('----')
```

```
a a
a b
a c
a d
----
b a
b b
```

Edit View Insert Cell Kernel Widgets Help

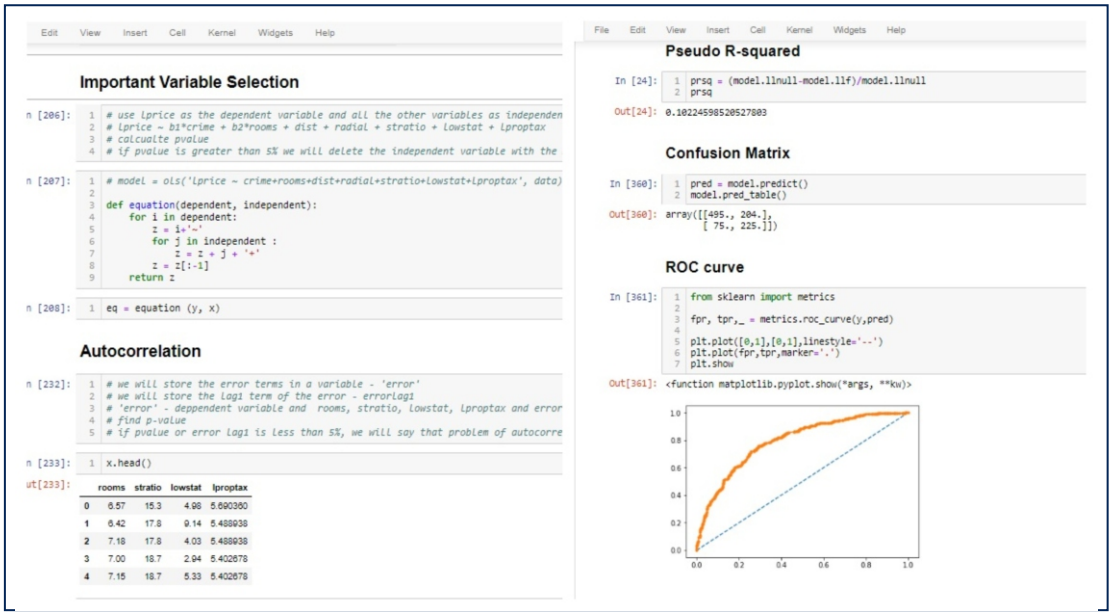

```
In [74]:
plt.scatter(y=sales,x=newspaper)
```

```
Out[74]: <matplotlib.collections.PathCollection at 0xbcf9910>
```



```
In [75]:
plt.scatter(y=sales,x=radio)
```

```
Out[75]: <matplotlib.collections.PathCollection at 0xb587130>
```



Edit View Insert Cell Kernel Widgets Help

Important Variable Selection

```
In [206]:
# use lprice as the dependent variable and all the other variables as independen
# lprice ~ b1*crime + b2*rooms + dist + radial + stratio + lowstat + lproptax
# calcualte pvalue
# if pvalue is greater than 5% we will delete the independent variable with the
```

```
In [207]:
# model = ols('lprice ~ crime+rooms+dist+radial+stratio+lowstat+lproptax', data)

def equation(dependent, independent):
    for i in dependent:
        z = i+'~'
        for j in independent :
            z = z + j + '+'
        z = z[:-1]
    return z
```

```
In [208]:
eq = equation (y, x)
```

Autocorrelation

```
In [232]:
# we will store the error terms in a variable - 'error'
# we will store the lag1 term of the error - errorlag1
# 'error' - dependent variable and  rooms, stratio, lowstat, lproptax and error
# find p-value
# if pvalue or error lag1 is less than 5%, we will say that problem of autocorre
```

```
In [233]:
x.head()
```

```
Out[233]:
```

	rooms	stratio	lowstat	lproptax
0	6.57	15.3	4.98	5.690360
1	6.42	17.8	9.14	5.488938
2	7.18	17.8	4.03	5.488938
3	7.00	18.7	2.94	5.402678
4	7.15	18.7	5.33	5.402678

File Edit View Insert Cell Kernel Widgets Help

Pseudo R-squared

```
In [24]:
prsq = (model.llnull-model.llf)/model.llnull
prsq
```

```
Out[24]: 0.10224598520527803
```

Confusion Matrix

```
In [360]:
pred = model.predict()
model.pred_table()
```

```
Out[360]: array([[495., 204.],
                 [ 75., 225.]])
```

ROC curve

```
In [361]:
from sklearn import metrics

fpr, tpr,_ = metrics.roc_curve(y,pred)

plt.plot([0,1],[0,1],linestyle='--')
plt.plot(fpr,tpr,marker='.')
plt.show
```

```
Out[361]: <function matplotlib.pyplot.show(*args, **kw)>
```

FREQUENTLY ASKED QUESTIONS

PREREQUISITE

Knowledge of Basic Excel

CERTIFICATE

Silver Certificate on successful completion of projects .
Gold Certification on passing a 2 hours MCQ based exam.

FEES

Rs.20000

DURATION

50+ hours

ABOUT THE TRAINER

Karan Aggarwal is one of India's leading trainers in Financial Modelling, Risk Modelling, Data Analytics and academic programs like Financial Risk Manager (FRM) & Actuarial Science. He has spearheaded several solution accelerators and spreadsheet-based prototypes in Risk and Analytics space. Karan has also authored a number of books on Advanced Excel, Statistical Modelling, Risk Modelling & Machine Learning. He is widely regarded for his problem-solving, thought leadership and intrapreneurship skills. His analytical mindset, solid fundamentals & the thirst to keep learning set him apart as a true authority in this field. Karan has also been awarded the Young Indian Entrepreneur Award by the Confederation Of Indian Industries in the year 2017.

PEAKS 2 TAILS

OUR TRAINEES WORK IN

DUPONT

LG
Life's Good

Manyavar
EARN YOUR RESPECT

XLRI
Xavier School of Management
For the greater good

TATA MOTORS

Uber

PRIONE
BUSINESS SERVICES PRIVATE LIMITED
A Catamaran & amazon Venture

LONE STAR COLLEGE
UNIVERSITY PARK
The Honors College

St. Xavier's College
NIHIL ULTRA

GPI
GODFREY PHILLIPS
INDIA LIMITED

SUN
PHARMA

moneytor

सेल SAIL
स्टील अथॉरिटी ऑफ इण्डिया लिमिटेड
STEEL AUTHORITY OF INDIA LIMITED

primarc

WNS
Extending Your Enterprise

OUR SERVICES

1 2 3 4

TRAINING

SOLUTIONING

CONSULTING

98 74 98 74 98