We completed Python : Part-1 and Part-2

You have 10 days left

Statistics class , pen and paper

Document and upload in LinkedIn :

Statistics

Part-1: EDA Exploratory data analysis

* Python EDA

Part-2: Leads To ML models

80% EDA 20% ML

EDA theory easy : Practical is tricky (1 week)

ML theory is tricky: Practical is easy

Linear regression algorithm: 1 week

Practical : 1min

Contents:

1. What is statistics
2. Data types
3. Levels of data
4. Population
5. Sample
6. Frequency table
7. Bar graph
8. Pie chart
9. Frequency distribution table
10. Histogram
11. Distribution plot
12. Central tendency
13. Mean-Median-mode
14. Data dispersion
15. Range-Absolute Mean deviation
16. Variance-Standard deviation
17. Skew ness of data
18. Right skewed – Left skewed
19. Normal distribution
20. Outlier analysis
21. Covariance
22. Correlation
23. Standardization
24. Normalization
25. Convert categorical to Numerical

Statistics:

Maths numbers tables graphs formulae

Analysis

Data display

Probability

HDFC bank has one problem , they want to implement loan defaulter analysis

They approached

They given the data also

1. Understand the data
2. Analyse the data
3. Patterns
4. Represent the data
5. Summary of the data

Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data

Analyse ====== Interpreting

Data types :

Int float string bool complex

1. Those which are having numbers

10 10.5 ======= numerical data

or

Quantitative data

1. Discrete data : integer data 10 20 30
2. Continuous data: 10.54321

How many kids : Discrete data

Cont. data: height 152.4 cm

Weight : 57.0

Temp : 98.4F

1. Those which are having English letters

Categorical data

or

Qualitative data

|  |  |
| --- | --- |
| Numerical data | Categorical data |
| Quantitative data | Qualitative data |
| Discrete: Roll numbers, Id | Yes or No |
| Conti. Ex: Height, temp, salary |  |

We divided data into 4 levels

* Nominal data --------- > categorical

Vijay alex Pankaj sahil manha

Hyd London Italy

* Ordinal data ---------- > categorical

You will see the relation

Flop average hit superhit blockbuster

Best Good bad worst

Low medium high

Nano Micro mini medium

A B C D

Yes NO

* Interval data ---------- > numerical

You will see some interval

Interval data has no zero scale

* Ratio level data -------- > numerical

Ratio level data have zero scale

Best example

Temperature in Hyd is 50 degrees Celsius ------------ 122F

Temperature in Blr is 25 degrees Celsius ------------- 77 F

Temp(Hyd)= 2\* Temp(Blr)

Temp(Hyd)/ Temp(Blr)= 2/1 122/77= 2

Kid weight=25kgd ========== 55 Father w= 50kgs================ Pounds=110

F/K= 50/25=2 110/55=2

Temperature has negative values ========= Confirm

Temperature has No zero scale : Ratio is failed ===== Interval level

Conclusion.

Interval level has No zero scale

Ratio level has zero scale

Population vs sample:

Population is generally has each and every observation

Sample is part of the populations

Ex: 1lakh rows of data you are assuming that it is a population

You want to select 20k observation that it is a sample

Population never exist ======= myth

The total population in india 140B : 100

It is very difficult to work on population

Analyst ========= which party will comes in to power

140B

1. You need more time
2. Require more resources
3. More Money

Two types of analysis are there

1. You will work on population, you will make conclusion on population

You will go to each and every home ======== on India

Descriptive statistics

1. You will work on sample , you will make conclusion on population

Hyd : 32 parts ==== > 100 3200

Blr : 32 parts==== 100

Infer ========

Inferential statistics

Open Intro Statistics

Ken black

Leonard root(25)

Khan academy

Links also I will provide ===== DS c