SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME :19CS407 DATA ANALYTICS WITH R
II YEAR /IV SEMESTER

Unit 1- Introduction

Topic : Data Analytics- Tools and Applications

DATA ANALYTICS/JEBAKUMAR IMMANUEL/CSE/SNSCE

Data Analytics Tools

- ✓ Microsoft Excel
- ✓ Python
- ✓ R
- ✓ Jupyter Notebook
- ✓ Apache Spark
- ✓ SAS
- ✓ Microsoft Power BI
- ✓ Tableau
- ✓ KNIME

Microsoft Excel

- **Type of tool:** Spreadsheet software.

- **Availability**: Commercial.

- **Mostly used for:** Data wrangling and reporting.

- **Pros:** Widely-used, with lots of useful functions and plug-ins.

- **Cons:** Cost, calculation errors, poor at handling big data.

Python

- **Type of tool:** Programming language.

- **Availability:** Open-source, with thousands of free libraries.

- **Used for:** Everything from data scraping to analysis and reporting.

- **Pros:** Easy to learn, highly versatile, widely-used.

- **Cons:** Memory intensive—doesn't execute as fast as some other languages.

R

- **Type of tool:** Programming language.

- **Availability:** Open-source.

- **Mostly used for:** Statistical analysis and data mining.

- **Pros:** Platform independent, highly compatible, lots of packages.

- **Cons:** Slower, less secure, and more complex to learn than Python.

Jupyter Notebook

- **Type of tool:** Interactive authoring software.

- **Availability:** Open-source.

- **Mostly used for:** Sharing code, creating tutorials, presenting work.

- **Pros:** Great for showcasing, language-independent.

- **Cons:** Not self-contained, nor great for collaboration.

Apache Spark

- **Type of tool:** Data processing framework.

- **Availability:** Open-source.

- **Mostly used for:** Big data processing, machine learning.

- **Pros:** Fast, dynamic, easy to use.

- **Cons:** No file management system, rigid user interface.

SAS

- **Type of tool:** Statistical software suite.

- **Availability:** Commercial.

- **Mostly used for:** Business intelligence, multivariate, and predictive analysis.

- **Pros:** Easily accessible, business-focused, good user support.

- **Cons:** High cost, poor graphical representation.

Microsoft Power BI

- **Type of tool:** Business analytics suite.

- **Availability:** Commercial software (with a free version available).

- **Mostly used for:** Everything from data visualization to predictive analytics.

- **Pros:** Great data connectivity, regular updates, good visualizations.

- **Cons:** Clunky user interface, rigid formulas, data limits (in the free version).

Tableau

- **Type of tool:** Data visualization tool.

- **Availability:** Commercial.

- **Mostly used for:** Creating data dashboards and worksheets.

- **Pros:** Great visualizations, speed, interactivity, mobile support.

- **Cons:** Poor version control, no data pre-processing.

KNIME

- **Type of tool:** Data integration platform.

- **Availability:** Open-source.

- **Mostly used for:** Data mining and machine learning.

- **Pros:** Open-source platform that is great for visually-driven programming.

- **Cons:** Lacks scalability, and technical expertise is needed for some functions.

Assessment 1

References

1. João Moreira, Andre Carvalho, Tomás Horvath – "A General Introduction to Data Analytics" – Wiley -2018

Thank You