INT217:INTRODUCTION TO DATA MANAGEMENT

L:2 T:0 P:2 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: apply the various techniques and functions over spreadsheet for getting various insights of data

CO2 :: practice the data representation methods like pivot table and power pivoting

CO3 :: determining the need of the graphical representation in the spreadsheet by using various graphs and charts

CO4 :: understand the concepts of multidimensional data and topics like OLAP, OLTP and data cube

CO5 :: study the ETL process using the SSIS tool and able to perform pre-processing for data analysis

CO6 :: describe effective data management is central to good research practice and identify the benefits of effective data management.

Unit I

Spreadsheet functions to organize data : cell reference styles, creating and working with formulas, text functions, date and time functions, lookup and reference functions, mathematical and statistical functions, information and volatile functions, logical and financial functions, formula auditing, error handling, string functions

Introduction to Microsoft Excel : UI Basics : introduction to UI basics, about excel, workbooks and worksheets, customizing excel, reference styles, number formatting, custom number formatting, conditional formatting, format as table

Unit II

Data representation and manipulation : filter, advanced filter for complex criterion, sorting and custom sorting, pivot table and pivot chart, power pivot, import data from different sources into power pivot, reducing file size in power pivoting, connect to multiple different external datasets, DAX functions

Unit III

Advanced graphing and charting : charts, combo charts, working with objects charts, dynamic charts and dynamic data source for charts print areas, views for a worksheet, various printing techniques

Data protection techniques : worksheet protection, protect specific range, workbook protection and encryption

Unit IV

Building a model : introduction to building a model, about model and assumptions, diagramming the flow with exercise, wireframing with exercise, preparing the data with exercise, linking up formulas with exercise, sensitivity analysis with exercise, case study: forecasting with a model

Unit V

Multidimensional data : ETL overview, extracting data, transformations, loading data, simple ETL processing, ETL tools, data sources and destinations, OLTP and OLAP, multidimensional data models, data cube, grouping sets in T-SQL

Unit VI

Data Processing using Tableau Prep : introduction to tableau prep, installing tableau prep, basics of data preparation, connecting to data, creating data flow, blend data sources, cross-database join, project task, tableau prep for business

List of Practicals / Experiments:

Introduction to spreadsheets

- basic terminology of excel
- spreadsheet environment
- object model of excel
- customizing excel

- reference styles
- number formatting
- custom number formatting
- conditional formatting
- format as table

ETL processing with SSIS

- SQL programming for data science
- creating SSIS projects
- data wrangling before the load
- uploading data
- handling errors during ETL
- data wrangling after the load
- testing and deploying of the project

References:

1. FUNDAMENTALS OF BUSINESS ANALYTICS by R.N. PRASAD, SEEMA ACHARYA, WILEY

2. EXCEL HACKS,2/ED TIPS & TOOLS FOR STREAMLINING YOUR SPREADSHEETS by DAVID, SHROFF/O'REILLY