BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

WORK INTEGRATED LEARNING PROGRAMMES

Part A: Content Design

Course Title	Introduction to Data Science
Course No	DSE ZG523
Credit Units	3
Last Revised by	
Version / Date	

Course Objectives

#	Course Objectives
1	Gain basic understanding of the role of Data Science in various scenarios in the real-world of business, industry and government
2	Appreciate the application of the concepts of Statistics, Linear Algebra and Graph Theory in Data Science/Machine learning etc.
3	Understand various roles and stages in a Data Science Project and ethical issues to be considered.
4	Explore the processes, tools and technologies for collection and analysis of structured and unstructured data

Text Books/References

ID	Text/Ref Book		
T1	Introducing Data Science by Cielen, Meysman and Ali		
T2	Storytelling with Data, A data visualization guide for business professionals, by Cole Nussbaumer Knaflic; Wiley		
Т3	The Art of Data Science by Roger D Peng and Elizabeth Matsui		
R1	Ethics and Data Science by DJ Patil, Hilary Mason, Mike Loukides		
R2	KDD, SEMMA and CRISP-DM: A Parallel Overview , Ana Azevedo and M.F. Santos , IADS-DM, 2008		
R3	An Introduction to Data Science by Jeffrey Stanton (ebook)		
R4	Python Data Science Handbook: Essential tools for working with data by Jake VanderPlas		

^{*} The above materials are reference only and are neither conclusive nor exhaustive. However, the student is advised to refer latest content from online sources or instructor supplied materials for more thorough understanding of the topics

Content Structure

1. Introduction to Data Science

- 1.1. Definition
- 1.2. Need of Data Science
- 1.3. Motivating Examples
- 1.4. Roles and responsibilities of a Data Scientist
- 1.5. Data Science vs BI
- 1.6. Data Science vs Statistics
- 1.7. Data Science Applications
- 1.8. Data Science Concerns

2. Data Science Process

- 2.1. Roles in a Data Science project
- 2.2. Setting expectations
- 2.3. Data Science methodology
 - 2.3.1. Business understanding,
 - 2.3.2.Data Requirements,
 - 2.3.3.Data Acquisition,
 - 2.3.4. Data Understanding,
 - 2.3.5. Data preparation,
 - 2.3.6. Modelling,
 - 2.3.7. Model Evaluation,
 - 2.3.8.Deployment and feedback.
- 2.4. Case Study
- 2.5. Data Science Proposal Samples
- 2.6. Data Science Proposal Evaluation
- 2.7. Data Science Proposal Review Guide

3. Data

- 3.1. Data quality
- 3.2. Types of Data
- 3.3. Data Formats
- 3.4. High dimensional data
- 3.5. Data representation
 - 3.5.1. Graphs and networks,
 - 3.5.2. Matrices, Vectors,
 - 3.5.3. Data Frames, list
 - 3.5.4. Libraries of Graph, Matrices and vectors
- 3.6. Data Models
 - 3.6.1.Model as expectation
 - 3.6.2. Comparing models to reality
 - 3.6.3. Reactions to Data
 - 3.6.4. Refining our expectations
- 3.7. Data Sampling
 - 3.7.1. Probability sampling
 - 3.7.2. Non-Probability sampling

4. Data Wrangling

- 4.1. Handling Numeric Data
- 4.2. Dealing with textual Data
- 4.3. Managing Categorical Attributes
- 4.4. Transforming Categorical to Numerical Values
- 4.5. Feature Engineering
- 4.6. Feature Selection

- 4.6.1. Curse of dimensionality
- 4.6.2. Dimensionality Reduction
 - 4.6.2.1. Data Correlation,
 - 4.6.2.2. PCA
- 4.6.3. Nonlinear Featurization

5. Data Analytics

- 5.1. Definitions
- 5.2. Types of data analytics
 - 5.2.1. Predictive, Descriptive, Prescriptive, Diagnostic
- 5.3. Analytics terminology
- 5.4. Data analytics methodologies
 - 5.4.1.CRISP-DM Methodology
 - 5.4.2.SEMMA
 - 5.4.3.BIG DATA LIFE CYCLE
 - 5.4.4.SMAM
 - 5.4.5.ASUM- DM
- 5.5. Applications

6. Data visualization

- 6.1. Need for visualization
- 6.2. Exploratory vs Explanatory Analysis
- 6.3. Tables, Axis based Visualization and Statistical Plots
- 6.4. The Data Visualization Design Process
- 6.5. Lessons in Data Visualization Design
- 6.6. Stories and Dashboards

7. Ethics for Data Science

- 7.1. Why Data science needs Ethics
- 7.2. History, Concept of informed consent
- 7.3. Being a data sceptic
- 7.4. Ethical guidelines for Data Scientist
- 7.5. Data Science concerns
- 7.6. Data Privacy and Legal aspects
- 7.7. Societal consequences
- 7.8. Ethics of data scraping and storage
- 7.9. Rightful use of data science

8. Storytelling with Data

- 8.1. The final deliverable
- 8.2. The Narrative report / presentation structure
- 8.3. Building narrative with Data
- 8.4. Effective storytelling

9. Review

Contents & Session delivery

1. Introduction to Data Science Definition Need of Data Science Motivating Examples Roles and responsibilities of a Data Scientist Data Science vs BI Data Science vs Statistics Data Science Pocess Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Acquisition, Data Understanding, Data Understanding, Modelling, Modelling, Model Evaluation, Data Science Proposal Samples Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Quality Types of Data Data Propasal High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling Non-Probability sampling 6. Data Wrangling R4 - Chapter 1, 5	Session	Topics to cover	Content Reference
Definition Need of Data Science Motivating Examples Roles and responsibilities of a Data Scientist Data Science vs B1 Data Science vs B1 Data Science Statistics Data Science Applications Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Acquisition, Data Acquisition, Data Acquisition, Data Acquisition, Data Proparation, Model Evaluation, Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Probability sampling Pon-Probability sampling Non-Probability sampling Pon-Probability sampling Non-Probability sampling Posa Wranugling Refore the Matrices and Science Propersions To Chapter 1, 5	(2 hrs)		
Need of Data Science Motivating Examples Roles and responsibilities of a Data Scientist Data Science vs BI Data Science Vapilications Data Science Applications Data Science Concerns 2. Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Acquisition, Data Understanding, Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data Formats High dimensional data Data representation Oraphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling	1.		T1 – Chapter 1
Motivating Examples Roles and responsibilities of a Data Scientist Data Science vs BI Data Science vs Statistics Data Science Process Data Science Concerns 2. Data Science Process Roles in a Data Science project Setting expectations Data Science enthodology Business understanding, Data Requirements, Data Acquisition, Data Preparation, Model Evaluation, Data Data Science Proposal Samples Data Science Proposal Samples Data Science Proposal Review Guide 4. Data Data Obata Obata Obata Obata Obata Proposal Review Guide 4. Data Obata Formats High dimensional data Data representation Ofarphs and networks, Matrices, Vectors, Data Models Data Models Nodel as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Ref-Chapter 1, 5			
 Roles and responsibilities of a Data Scientist Data Science vs BI Data Science vs Statistics Data Science Applications Data Science Concerns 2. Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Requirements, Data Acquisition, Data Understanding, Model Evaluation, Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Evaluation Data Formats High dimensional data Data Formats Hitps://www.researchgate.net/p ublication/319998246 Sampling Methods in Research Method ology How to Choose a Sampl ing Technique for Research/lin k/59c5f8c2a6fdccc719164f0b/download 5. Data Sampling Probability sampling Probability sampling Non-Probability sampling 6. Data Wrangling R4 - Chapter 1, 5 		Need of Data Science	
Data Science vs BI Data Science Applications Data Science Concerns 2. Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Data Science methodology Data Requirements, Data Acquisition, Data Acquisition, Data Understanding, Data Understanding, Data Preparation, Modelling, Model Evaluation, Data Science Proposal Samples Data Science Proposal Review Guide 4. Data Data Science Proposal Review Guide 4. Data Data Science Proposal Review Guide 4. Data Data Framas Data Framas Data Frames, list Data Frames, list Data Models Model a expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Ref-Chapter I, 5		 Motivating Examples 	
Data Science vs Statistics Data Science Applications Data Science Concerns Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Data Science methodology Data Requirements, Data Requirements, Data Acquisition, Modelling, Model Evaluation, Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Evaluation Data Science Proposal Review Guide Data Science Proposal Review Guide Data Formats High dimensional data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Reflining our expectations Data Sampling Probability sampling Non-Probability sampling R4 - Chapter 1, 5		 Roles and responsibilities of a Data Scientist 	
Data Science Applications Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Acquirements, Data Acquisition, Data Understanding, Modelling, Modelling, Model Evaluation, Data Science Proposal Samples Data Science Proposal Samples Data Science Proposal Samples Data Science Proposal Review Guide 4. Data Data Data Graph, Matrices and vectors Data Frames, list Data Frames, list Data Frames, list Data Frames, list Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling R4 - Chapter 1, 5		Data Science vs BI	
Data Science Concerns 2. Data Science Process Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Requirements, Data Acquisition, Data Understanding, Data preparation, Model Evaluation, Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Samples Data Science Proposal Review Guide 4. Data Data Formats High dimensional data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Models Data Models O Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Non-Probability sampling R4 - Chapter 1, 5		 Data Science vs Statistics 	
2. Data Science Process • Roles in a Data Science project • Setting expectations • Data Science methodology • Business understanding, • Data Requirements, • Data Adquisition, • Data Understanding, • Model Evaluation, • Deployment and feedback. • Case Study • Data Science Proposal Samples • Data Science Proposal Evaluation • Data Science Proposal Review Guide 4. Data • Data Quality • Types of Data • Data Formats • High dimensional data • Data Formats • High dimensional data • Data representation • Graphs and networks, • Matrices, Vectors, • Data Models • Model as expectation • Model as expectation • Comparing models to reality • Reactions to Data • Refining our expectations • Data Sampling • Probability sampling • Non-Probability sampling • Non-Probability sampling • Non-Probability sampling • Non-Probability sampling • R4 - Chapter 1, 5		Data Science Applications	
Roles in a Data Science project Setting expectations Data Science methodology Business understanding, Data Requirements, Data Acquisition, Data Understanding, Data preparation, Modelling, Model Evaluation, Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling Net-Chapter 1, 5		Data Science Concerns	
Setting expectations Data Science methodology Business understanding, Data Requirements, Data Acquisition, Data Understanding, Data Understanding, Data Popparation, Modelling, Model Evaluation, Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Pormats High dimensional data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Nat - Chapter 1, 5	2.	Data Science Process	T1 - Chapter 2
Data Science methodology		Roles in a Data Science project	
Data Science methodology		Setting expectations	
Business understanding, Data Requirements, Data Acquisition, Data Understanding, Data preparation, Modelling, Data peployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling R4 - Chapter 1, 5		7 2	
O Data Requirements, O Data Acquisition, O Data Understanding, O Data preparation, O Modelling, Model Evaluation, O Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Guality T1 - Chapter 1 T3- Chapter 1 T3- Chapter 5 Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Science Proposal Evaluation T1 - Chapter 1 T3- Chapter 5 T3- Chapter 1 T3- Chapte			
O Data Acquisition, O Data Understanding, O Data preparation, Modelling, Model Evaluation, Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 5 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T3 - Chapter 5 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T1 - Chapter 1 T3 - Chapter 1 T4 - Chapter 1 T4 - Chapter 1 T3 - Chapter 1 T4 - Chapter 1 T4 - Chapter 1 T4 - Chapter 1 T4		_	
O Data Understanding, O Data preparation, O Modelling, O Model Evaluation, O Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation O Graphs and networks, O Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models O Model as expectation O Comparing models to reality Reactions to Data Refining our expectations Data Sampling Non-Probability sampling At - Chapter 1, 5		_	
O Modelling, O Model Evaluation, O Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Data Quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Bata Chapter 1 T1 - Chapter 1 T3- Chapter 5 https://www.researchgate.net/publication/319998246 Sampling Methods in Research Method ology How to Choose a Sampl ing Technique for Research/lin k/59c5f8c2a6fdccc719164f0b/do wnload R4 - Chapter 1, 5			
O Modelling, O Model Evaluation, O Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Data Quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Bata Chapter 1 T1 - Chapter 1 T3- Chapter 5 https://www.researchgate.net/publication/319998246 Sampling Methods in Research Method ology How to Choose a Sampl ing Technique for Research/lin k/59c5f8c2a6fdccc719164f0b/do wnload R4 - Chapter 1, 5	3		
O Deployment and feedback. Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Non-Probability sampling Ref- Chapter 1, 5	٥.		
Case Study Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Data Quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Pata Sampling Probability sampling Non-Probability sampling Non-Probability sampling Non-Probability sampling 6. Data Wrangling Pata Science Proposal Evaluation TI - Chapter 1 T3- Chapter 5 T3- Chapter 5 Methods in Research Method ology How to Choose a Sampling Methods in Research Method ology How to Choose a Sampling k/59c5f8c2a6fdccc719164f0b/do wnload Refining our expectations Pata Sampling Non-Probability sampling Non-Probability sampling R4 - Chapter 1, 5		 Model Evaluation, 	
Data Science Proposal Samples Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling 6. Data Wrangling Propagation T1 - Chapter 1 T3- Chapter 5 T3- Chapter 5 T3- Chapter 5 Methods in Research Method ology How to Choose a Sampling Methods in Research Method ology How to Choose a Sampling winload Propagation of Research Method ology How to Choose a Sampling winload Refining our expectations Probability sampling Non-Probability sampling R4 - Chapter 1, 5		 Deployment and feedback. 	
Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Non-Probability sampling 6. Data Wrangling Pata Science Proposal Review Guide T1 - Chapter 1 T3- Chapter 5 T3- Chapter 5 T3- Chapter 5 Attps://www.researchgate.net/publication/319998246 Sampling Methods in Research Method ology How to Choose a Sampling		Case Study	
Data Science Proposal Evaluation Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Non-Probability sampling 6. Data Wrangling Pata Science Proposal Review Guide T1 - Chapter 1 T3- Chapter 5 T3- Chapter 5 T3- Chapter 5 Attps://www.researchgate.net/publication/319998246 Sampling Methods in Research Method ology How to Choose a Sampling		1	
Data Science Proposal Review Guide 4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling Ref - Chapter 1 T3 - Chapter 1 T3 - Chapter 5 T3 - Chapter 5 https://www.researchgate.net/p ublication/319998246 Sampling Methods in Research Method ology How to Choose a Sampl ing Technique for Research/lin k/59c5f8c2a6fdccc719164f0b/do wnload			
4. Data Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Bata Chapter 1 T3- Chapter 5 T3- Chapter 5 https://www.researchgate.net/p ublication/319998246 Sampling Methods in Research Method ology How to Choose a Sampl ing Technique for Research/lin k/59c5f8c2a6fdccc719164f0b/do wnload Refining our expectations Data Sampling R4 - Chapter 1, 5		_	
 Data quality Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		•	
 Types of Data Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 	4.		T1 - Chapter 1
 Data Formats High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		_ · ·	TTO C1
 High dimensional data Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		1	T3- Chapter 5
 Data representation Graphs and networks, Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 			
Sampling O Graphs and networks, O Matrices, Vectors, O Data Frames, list O Libraries of Graph, Matrices and vectors Data Models O Model as expectation O Comparing models to reality O Reactions to Data O Refining our expectations O Data Sampling O Probability sampling O Non-Probability sampling O Reactions O Data Wrangling O Reference in the bubble of the sampling of the bubble of the bub		 High dimensional data 	https://www.wasaawah.gata.wat/g
 Matrices, Vectors, Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		 Data representation 	
 Data Frames, list Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		<u> </u>	
 Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5		I	Methods in Research Method
 Libraries of Graph, Matrices and vectors Data Models Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 	5	1	ology How to Choose a Sampl
 Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Another the Model of Market Parameters (Market Parameters) Model as expectation with a winload Winload Winload 		<u> </u>	ing Technique for Research/lin
 Model as expectation Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling A Data Wrangling Bata Wrangling R4 - Chapter 1, 5 			k/59c5f8c2a6fdccc719164f0b/do
 Comparing models to reality Reactions to Data Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		_	wnload
 Refining our expectations Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 			
 Data Sampling Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 			
 Probability sampling Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 			
 Non-Probability sampling Data Wrangling R4 - Chapter 1, 5 		1 0	
6. Data Wrangling R4 - Chapter 1, 5		· · · · ·	
		i	D4 G1 + 1.5
I I II II AT THE AT THE STATE OF THE STATE O	6.		_
Handling Numeric Data http://www.feat.engineering Class rests.		<u> </u>	
Dealing with textual Data Class notes			Class notes
Managing Categorical Attributes			
• Transforming Categorical to Numerical	7.		
Values			
Feature Engineering			
Feature Selection			
Curse of dimensionality		1	
O Dimensionality Reduction	R		
Data Correlation,	0.	· ·	
• PCA			
Nonlinear Featurization		 Nonlinear Featurization 	

Session(2 hrs)	Topics to cover	Content Reference
9.	Data Analytics	R2
10.	 Definitions Types of data analytics Predictive, Descriptive, Prescriptive, Diagnostic Analytics terminology Data analytics - methodologies CRISP-DM Methodology SEMMA BIG DATA LIFE CYCLE SMAM ASUM- DM Applications 	
11.	 Data visualization Need for visualization Exploratory vs Explanatory Analysis 	T2
12.	 Tables , Axis based Visualization and Statistical Plots The Data Visualization Design Process Lessons in Data Visualization Design Stories and Dashboards 	
13.	 Ethics for Data Science Why Data science needs Ethics History, Concept of informed consent Being a data sceptic Ethical guidelines for Data Scientist Data Science concerns 	R1 https://hbr.org/2013/04 /the-hidden-biases-in- big-data https://www.oreilly.co m/data/free/files/being-
	 Data Privacy and Legal aspects Societal consequences Ethics of data scraping and storage Rightful use of data science 	<u>a-data-skeptic.pdf</u>
15.	 Storytelling with Data The final deliverable The Narrative - report / presentation structure Building narrative with Data Effective storytelling 	T2
16.	Review	

Evaluation Scheme:

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Wt.	Date/Deadline*
EC-1A	Quiz-I (Pre-Mid / 20MCQ)	Online	3 days open	5%	
EC-1B	Quiz-II (Post-Mid / 20MCQ)	Online	3 days open	5%	
EC-1C	Assignment	Take-home	3 weeks	10%	
EC-2R	Mid-Semester Regular	Closed Book	1.5 hours	30%	
EC-2M	Mid-Sem Makeup	Closed Book	1.5 hours	30%	
EC-3R	Comprehensive Exam	Open Book	2.5 hours	50%	
EC-2M	Compre Makeup	Open Book	2.5 hours	50%	

Notes:

- → The release dates of Quiz-1/2 and assignments will be 3 days (for Quiz) and 3 weeks (for assignments) before the completion/submission deadline
- → Deadlines will NOT be extended for whatever reason and the student is requested not to wait for the deadline to start working on Quiz/Assignment
- → Syllabus for Quiz-I: Sessions: 1 to 4 / Quiz-II (all Sessions)
- → Syllabus for Assignment: Hands-on Python-based Exercise (real-world problem, for individual group of 3 students) / Group formation will be announced before Assignment release
- → All Quiz/Assignments will be released and to be answered/submitted in Canvas LMS
- → Syllabus for Mid-Semester Test (Closed Book): Topics in Session Nos. 1 to 8
- → Syllabus for Comprehensive Exam (Open Book): All topics (Session Nos. 1 to 16)
- → The student is strictly advised to stick to regular schedule of Mid-Sem and Compre examinations, and Makeup examinations will be only for those students with business-related absence/health related issues.
- → Strictly NO MAKEUPS for Quiz and Assignments and all submissions after the above stated deadlines will not be considered/evaluated.
- → All students should conform to BITS students' ethical code-of-conduct and all assignments will be subjected to plagiarism check, and if violated will be subject to disciplinary action apart from nullifying all the marks/grades assigned.

Important links and information:

<u>Canvas LMS</u>: All materials/announcements/discussions forums/Online Quizs/Assignment submissions will be via Canvas LMS portal. Students are expected to monitor this portal regularly for any content or announcements.

<u>Contact sessions:</u> Students should attend the online lectures as per the schedule provided in the Course Handout (posted on Canvas LMS)

Evaluation Guidelines:

- 1. EC-1 consists of 2 Quizzes and 1 Assignments. Students will attempt them through the course pages on the Canvas portal. Announcements will be made on the portal, in a timely manner.
- 2. For Closed Book tests: No books or reference material of any kind will be permitted.
- 3. For Open Book exams: Use of books and any printed / written reference material (filed or bound) is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
- 4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam which will be made available on the Elearn portal. The Make-Up Test/Exam will be conducted only at selected exam centres.

It shall be the responsibility of the individual student to be regular in attending the contact-session schedule as given in the course handout, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout