Course Code	PAD21102J	Course Name	DATA ANALYSIS FUNDAMENTALS	Course Category	C	Professional Core Course	L	T	P	C
							4	0	4	6

Pre-requisite Courses	Nil	Co-requisite Courses	Nil	Progressive Courses	Nil
Course Offering Department	Computer Applications	Data Book / Codes/Standards	Nil		

Course Learning Rationale (CLR):	The purpose of learning this course is to:	Learning			Program Learning Outcomes (PLO)														
CLR-1 :	Gather extensive knowledge in Data analysis and fundamental techniques	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-2 :	Improve and understand effective python programming language	Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	Fundamental Knowledge	Application of Concepts	Link with Related Disciplines	Procedural Knowledge	Skills in Specialization	Ability to Utilize Knowledge	Skills in Modeling	Analyze, Interpret Data	Investigative Skills	Problem Solving Skills	Communication Skills	Analytical Skills	ICT Skills	Professional Behavior	Life Long Learning
CLR-3 :	Strengthen the knowledge on numpy and Pandas tools.																		
CLR-4 :	Improve the problem-solving quality using data structure techniques																		
CLR-5 :	Gather extensive knowledge in Data Structures																		
CLR-6 :	Strengthen the knowledge in algorithms.																		
Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:																		
CLO-1 :	Apply basic mathematical concepts in Data Analysis	2	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-2 :	Work with powerful framework	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-3 :	Deal with Numpy and Pandas tools	3	75	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-4 :	Analyze various types data structure techniques	3	85	80	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-5 :	Apply various network models in deep learning	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-6 :	Apply data sampling techniques	3	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M

Duration (hour)		24	24	24	24	24
S-1	SLO-1	*Introduction to Data Science and Python Programming foundations*	*Advanced Python*	Pandas for Data Analysis	NumPy Basics and Manipulation	Statistical foundation for Data Science
	SLO-2	What is data science – Use cases of data science – Tools and programming languages	Generators – Generators expressions – Brief tour of standard library	Built-in styles – sharing styles – other options – fun stuff – export to excel	NumPy basic operations – Universal functions	**Data sampling terms – sampling bias – simple random sampling –**
S-2	SLO-1	Installation of Python –Using the Python interpreter – Invoking the interpreter – Argument passing – Interactive mode	Operations system interface	extensibility – options and settings – getting and setting options – environment	**indexing and slicing – iterating over arrays**	systematic random – sampling – stratified sampling
	SLO-2	The interpreter and its environment – Source coding	File wildcards– Command line arguments –	Frequently used options – available option	Array reshape – Image as bumpy arrays	**Non-probability sampling – Gaussian distribution – Inferential statistics and Hypothesis testing**
S-3	SLO-1	Informal introduction to Python – Numbers – Strings – Lists – first steps towards the programming	Error output redirection and program termination	Number formatting – Unicode formatting –	**Views and bumpy arrays**	**Hypothesis testing applied – T-Test – kurtosis and skewness**
	SLO-2	More control flow tools – if statements – Range function – Break and Continue statements – and else clause on loops – Pass statements	String pattern matching – Mathematics	Table schema display – enhancing performance	Creating array deep copies	**Correlation and autocorrelation**
S-4	SLO-1	Defining the functions – more on defining functions	Internet access – Date and times – Data compression	Python (writing c extensions for pandas) –using Numba	Understanding and applying index masks – structured arrays	**Introduction to linear – Regression – model fitting**
	SLO-2	Default arguments values – Keyword arguments – Special parameters Positional-or-keyword arguments –Positional(only parameters) – Keywords(only argument)	Performance measurement – Quality control	Expression evaluation via eval() – scaling to large datasets – load less data	**Understanding array broadcasting**	**Descriptive and Inferential Statistics – population vs sample –**
S 5 – S 8	SLO-1	**Implement a Python program to calculate GCD of two numbers**		**Solve algorithmic problems by program using different problem-solving strategies**		**Demonstrate handling of missing data**

		Implement a Python program to calculate the square root of a number by Newton's method	**Demonstrate use of and query()**		**Perform Linux administration task using Python**	
S9	SLO-1	Function examples – Arbitrary arguments list –Unpacking argument lists	Batteries included – brief tour of the Standard Library	use efficient data types – use chunking – use other libraries	Iterating on data frame contents	**Probability Vs Non-Probability sampling Mean/Median/Model**
	SLO-2	Lambda expressions – Documentation strings – Function annotations	Output formatting – Templating	migrating	**Exporting a data frame – sorting**	**QR – variance – The One-Sample T-Test – Independent and paired T-Test**
S10	SLO-1	Data structures – more on Lists – using Lists as Stacks – using List as Queues – List comprehensions – the del statement	Working with binary data record layouts	Intereraction with script sparse	**Handling missing data**	**testing – Hypotheses With T-Tests – loading and analysing A Skewed dataset**
	SLO-2	Tuples and Sequences – Sets – Dictionaries – Looping techniques –	Muti-threading – Logging	Data frame memory usage –	Grouping with index – merging with index	**measuring skewness and kurtosis – ANOVA test – interpretation of ANOVA**
S11	SLO-1	more on conditions – Comparing sequences and other types	weak references – Tools for working with lists –	Using if/truth statements with Pandas – nan – differences with NumPy	**Data-type descriptions – basic indexing (slicing)**	two way ANOVA discrete vs continuous distribution – pdf and cdf
	SLO-2	Modules – executing modules as scripts – the module search path – Complied python files	Decimal floating-point arithmetic –	Thread-salty – byte ordering issues – selection – multi-indexing	Memory layout of ndarray	Binomial distribution – interval estimation
S12	SLO-1	Standard modules – the dir() function packages importing * from a package – Intra- package references – Packages in multiple directories	Virtual environments and packages –	Missing data – grouping – timeseries	**Universal functions for arrays**	**Point and interval estimation Bayesian probability and Statistical Inference**
	SLO-1	Input and output – fancier output formatting	Introduction – creating virtual environments –	Merge – plotting – data in/out – computation –	Ndarray attributes – ndarray metds	Bayes theorem in machine learning – frequentist and subjective Probability

S 13 – S 16	SLO-1	**Implement a Python program to calculate the exponentiation of a number** **Implement a Python program to calculate the maximum from a list of numbers**	**Implement function overloading with different function signatures.**	**Search content using regular expression library in Python**	**Demonstrate aggregation**	**Demonstrate hierarchical indexing**
S17	SLO-1	Input and output – fancier output formatting	Managing packages with pip	time deltas – aliasing axis names – indexing	**Array shape manipulation**	Probability distribution – ingredients of – Bayesian statistics
	SLO-2	Methods of file objects saving structured data with json –	interactive input editing and history editing –	iteration – binary operator functions – function application	**Array indexing**	**Bayesian methods – Bayesian concepts in ml modelling – prior knowledge distribution**
S18	SLO-1	errors and expectations – Syntax errors – Exceptions – handling Exceptions – Raising exceptions – User defined exceptions	alternatives to the interactive Interpreter –	groupby & window – computations/descriptive stats	Operations on two or more arrays	Bayesian analysis approach –
	SLO-2	Defining clean-up actions – predefined clean up actions classes	Floating point arithmetic: issues and limitations –	reindexing/selection/label manipulation – Missing data handling	Shape functions – set operations	Bayesian learning – Bayesian – model types – probabilistic programming
S19	SLO-1	names and objects – Python scopes and namespaces	Representation error	Reshaping – voting – combining/joining/merging – time series-related	Array construction using index tricks	Modelling with PyMC – Bayesian data – analysis process
	SLO-2	scopes and namespace example – a first look at classes – Class definition syntax	appendix – interactive mode -	Metadata – plotting- sparse accessor	Two dimensional functions –	Bayesian data analysis with PyMC – Bayesian computation methods
S20	SLO-1	Class objects – Instance objects – Method objects – class instance variables	error handling	serialization/io/conversion – pandas array – period – panel	Method of array scalars	Markov chain simulation
	SLO-2	Random remarks – Inheritance – Multiple Inheritance – private	executable python scripts	Index – window – groupby – style – plotting	**Special attribute and methods recognized by NumPy**	**Implementing** Markov chain simulation – finding posterior modes

		variables – Odds and ends Iterators				
S 21 - S 24	SLO-1	**Implement a Python program to calculate the most frequent words in a text read from a file**	**Implement concept of class, instances and inheritance**	**implement matrix multiplication using multi-threading in Python**	**Demonstrate indexing and sorting**	**Demonstrate usage of pivot table**

Learning Resources	1. Python Crash Course, 2nd Edition, By Eric Matthes, May 2019 2. NumPy Essentials, By Leo Chin and Tanmay Dutta, April 2016 3. Hands-On Data Analysis, By Stefanie Molin, July 2019 4. The Python Workbook by Ben Stephenson, Springer, 2014 **5.** Hands-OnData Analysis, By Stefanie Molin, July 2

Learning Assessment											
Level	**Bloom's Level of Thinking**	**Continuous Learning Assessment (50% weightage)**								**Final Examination (50% weightage)**	
		CLA – 1 (10%)		**CLA – 2 (10%)**		**CLA – 3 (20%)**		**CLA – 4 (10%)#**			
		Theory	**Practice**	**Theory**	**Practice**	**Theory**	**Practice**	**Theory**	**Practice**	**Theory**	**Practice**
Level 1	Remember Understand	20%	20%	15%	15%	15%	15%	15%	15%	20%	20%
Level 2	Apply Analyze	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
Level 3	Evaluate Create	10%	10%	15%	15%	15%	15%	15%	15%	10%	10%
	Total	100 %		100 %		100 %		100 %		100 %	

CLA – 4 can be from any combination of these: Assignments, Seminars, Tech Talks, Mini-Projects, Case-Studies, Self-Study, MOOCs, Certifications, Conf. Paper etc.,

Course Designers		
Experts from Industry	**Experts from Higher Technical Institutions**	**Internal Experts**
Mr.G.Muruganandam, Group Project Manager, HCL Technologies, Chennai	Dr.Muthu, Professor, Loyola College, Chennai	*Mr.J.Venkata Subramanian, Assistant Professor*
Mr.M. Hemachandar, Tech Lead, Wipro Limited, Chennai	Dr.Vincent, Associate Professor, VIT	