PG Certificate Program

Data Science and Business Analytics

Session 1-2: Introduction to Data Science and Business Analytics

Instructor: Prof. Mahima Gupta

Instructor(s) email id: mahima.gupta@iimamritsar.ac.in

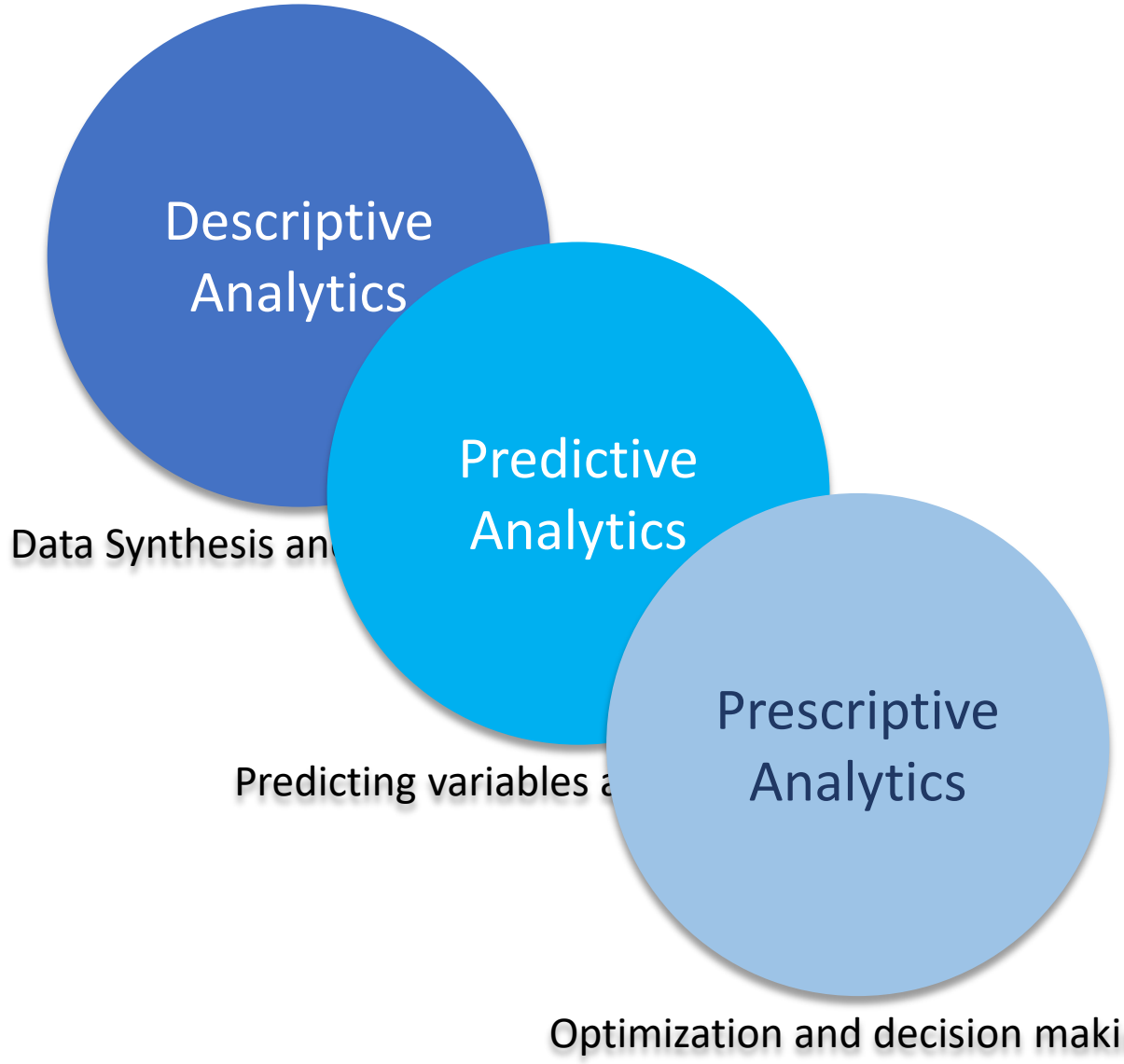
Session Objectives

O1: Data Driven Decision Making: An Illustration

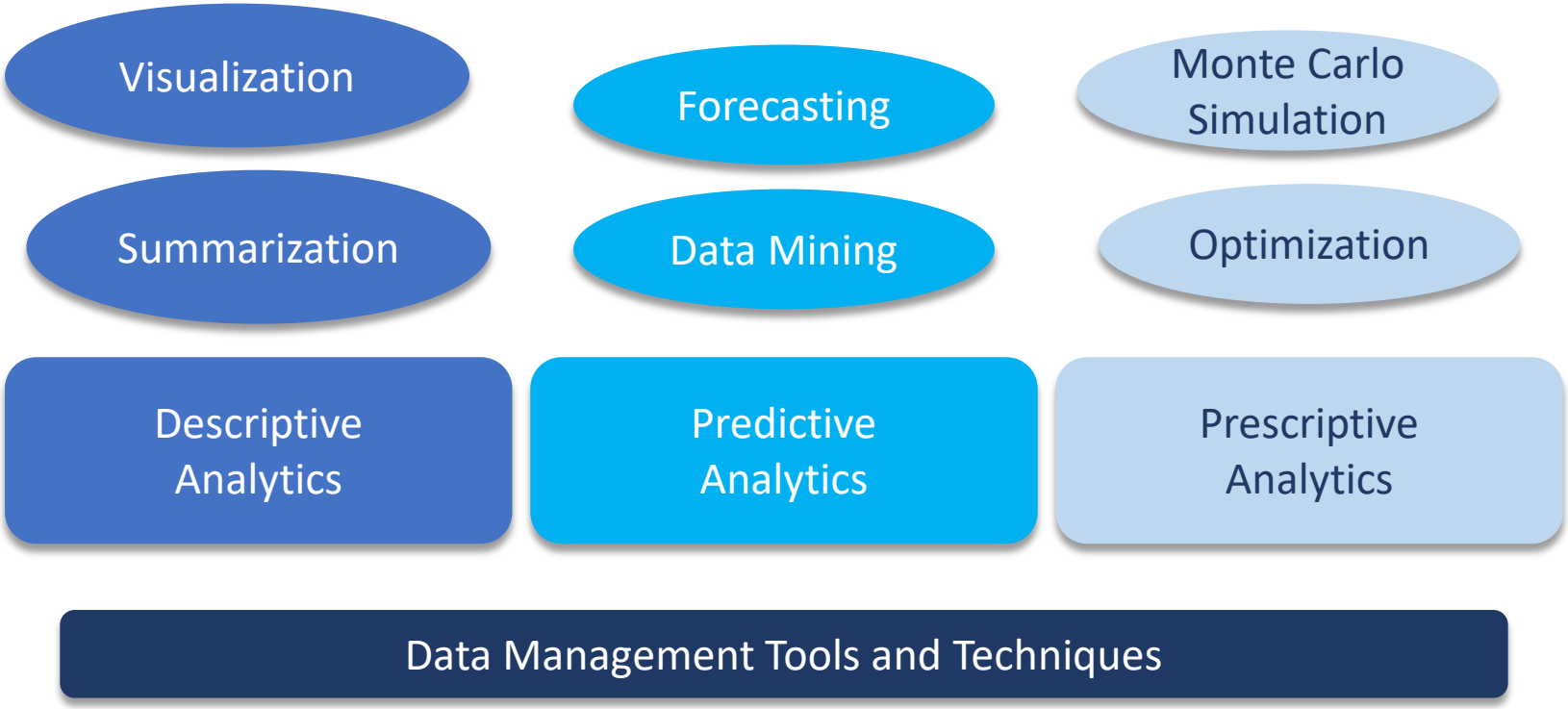
O2: Components of Analytics

O3: CRISP Data Mining Framework

Components of Analytics



Business Analytics Domain



Understanding with Data

INDUSTRY-WISE URBAN / HIGHWAY & RURAL RETAIL OUTLETS AS ON 01.10.2022

Statewise Sales Data

Some Examples

Descriptive Analytics
What is the total number of Retail Outlets (ROs) in rural or urban areas?
Is there a relationship between the number of ROs at urban locations by X and Y?
Does X focus less on highway segment than its competitors?
Does X has more or less similar rural, urban and highway mix across the states?

ROs	Urban	Highway	Rural
X	3559	7810	4370
Others	11798	30714	57730

Some Examples

Predictive Analytics
Predict the sales in a state by the number of ROs in highways, rural and urban locations.
Predict the sales in a state by the ROs mix & Competitors' ROs mix.
Predict whether a customer will churn a product/ brand based on his buying behavior and interaction with the company (such as Loyalty plan etc).
Predict whether a customer would like a product or respond to particular advertisement medium based on the customer profile and product characteristics.

Data				
	URBAN	HIGH WAY	RURAL	Sales
Andhra Pradesh	251	465	322	144.0
Arunachal Pradesh	0	1	0	5.7
Assam	20	59	31	49.7
Bihar	93	280	180	121.1
Chandigarh	11	0	0	19.1
Chhattisgarh	98	167	139	78.7
Dadra & Nagar Haveli	3	8	0	2.6
Daman & Diu	5	6	0	1.9

Sales = 3.41 +0.239*Urban + 0.4961*Highway-0.2*Rural

Some Examples

Prescriptive Analytics
Out of the possible n locations, which locations should be selected for the retail outlets?
Given the market risk and currency risk, we should buy from which suppliers.
What should be the right advertisement mix for the company?
What should be the right mix of offline, online centers for omni-channel retailing?

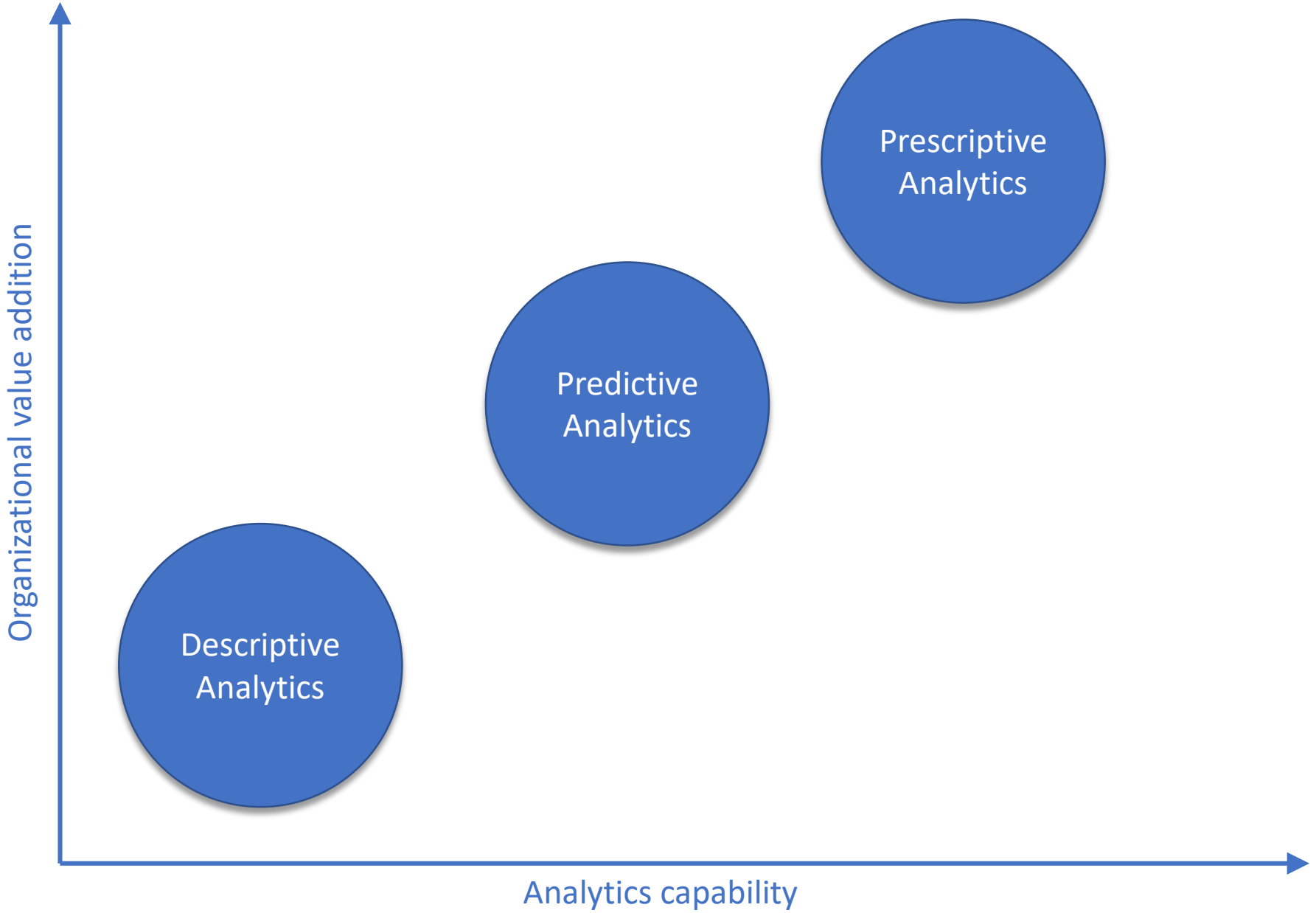
	Loc1	Loc2	Loc3	Loc4
Monthly Cost of Operations	25,000	10,000	15,000	30,000
Expected Sales	40,000	50,000	60,000	1,20,000
Monthly_Budget_Available	60,000			
Number of outlets	3			

Constraints

	Loc1		Loc2	Loc3	Loc4
	No		Yes	Yes	Yes
Maximize_Sales		230000			
Budget_Constraint		55000	<=	60000	
Total number of outlets		3	<=	3	

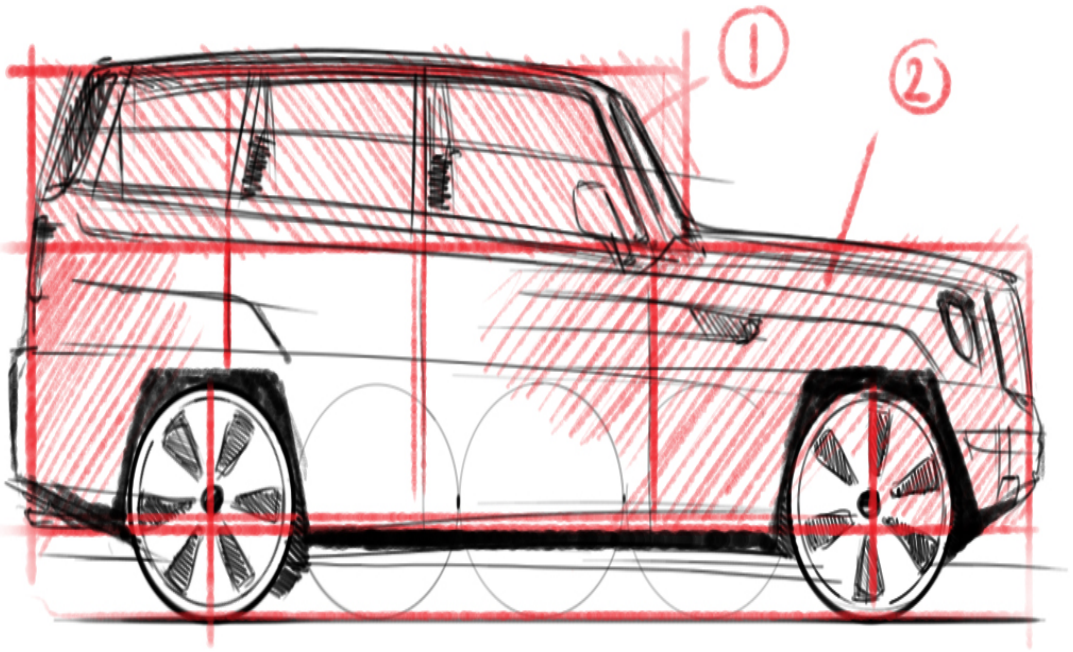
The Questions that are Asked…

Descriptive Analytics	Predictive Analytics	Prescriptive Analytics
What Happened?	What Will Happen?	What Action to Take?


Organizational value addition
Descriptive Analytics
Predictive Analytics
Prescriptive Analytics
Analytics capability

Picking Your Spot on Analytics

TARGET



Industry Sector	Sample Analytical problems	Data Sources
Manufacturing	• Supply Chain Analytics • Quality and process improvement • Revenue and cost management • Warranty Analytics	• Procurement , sales and production data • Warranty and after sales service • Commodity Price Data • Manufacturing Data • Macroeconomic Data
Retail	• Assortment Planning • Promotion Planning • Demand forecasting • Market Basket Analysis • Customer Segmentation	• Price data • Demand data at SKU and at category level • SKU level sales data with and without promotions • Planogram • Customer demographics data • Point of sales data • Loyalty program data
Healthcare	• Clinical care • Hospitality related data	• All patient care related data • Hospitality related data • Patient feedback data

Industry Sector	Sample Analytical problems	Data Sources
Service	• Demand forecasting • NPS Optimization • Service Quality Analysis • Customer Segmentation • Promotion	• Transactional and feedback data • Pricing and demand data • Promotional data
Banking & Finance	• Assortment Planning • Promotion Planning • Demand forecasting • Market Basket Analysis • Customer Segmentation	• Customer transactional data • Loan originating data • Credit scoring data
IT and ITES(IT enabling Services)	• Demand for Analytics Services • Software Development Cycle time	• Customer interaction and market research data • Internal product development data

**Primary sources of data and secondary sources to be used in solving these analytical problems

FINDING YOUR SPOT ON ANALYTCIS

EXERCISE: Churn Dataset

DESCRIPTIVE ANALYTICS	PREDICTIVE ANALYTICS	PRESCRIPTIVE ANALYTICS

Software

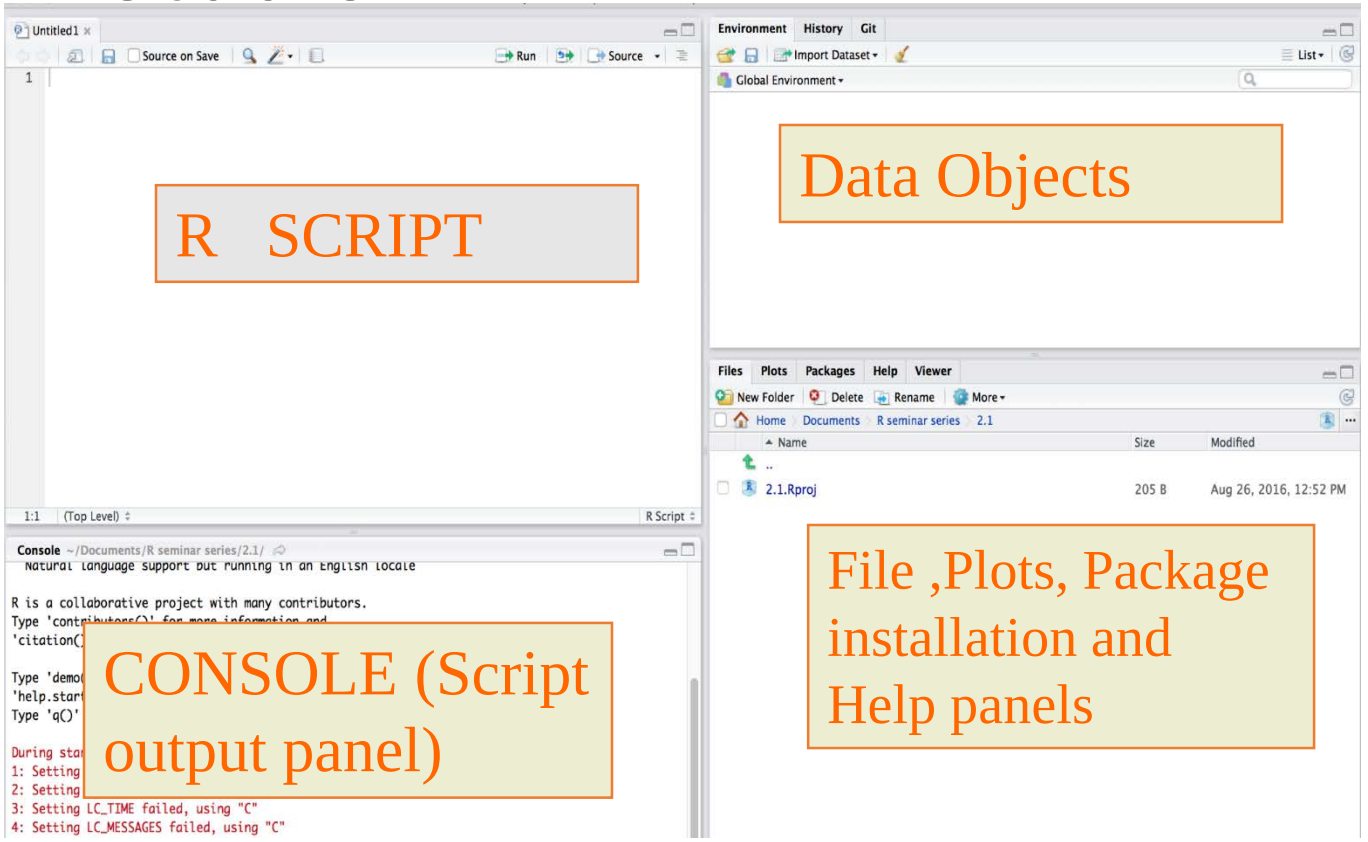
- Microsoft Excel or other spreadsheet programs like Google Sheets
- Proprietary Statistical Software: SAS, Stata or SPSS

- Limitations:
- Excel cannot handle datasets above a certain size.
- Reproducing previously conducted analyses on new datasets is challenging.
- Programs like SAS were developed for very specific uses.
- Don’t have a large community of contributors constantly adding new tools.

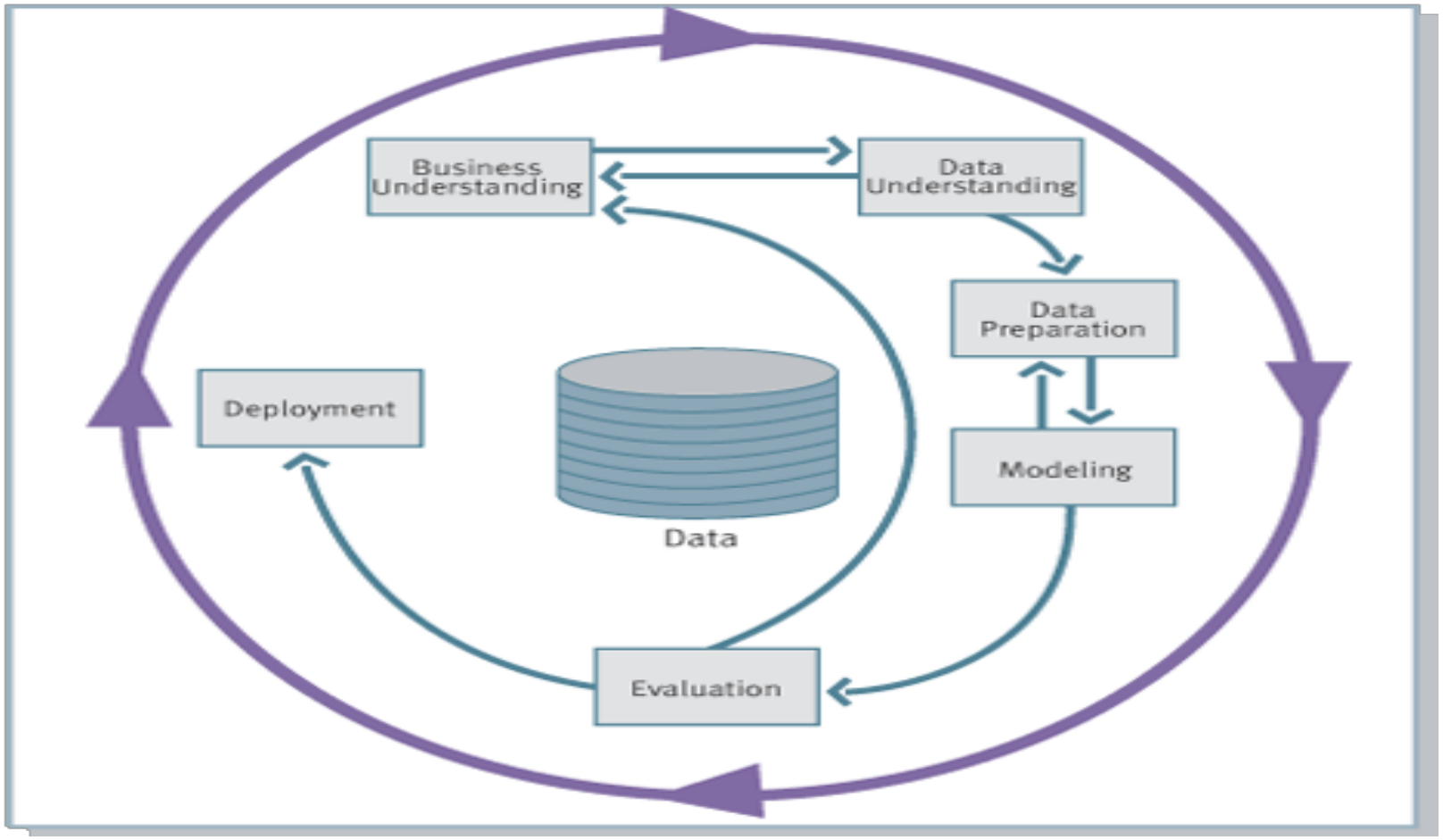
- R or Python

  - Both are free and open source, and were developed in the early 1990s.
  - R for statistical analysis and Python as a general-purpose programming language.
  - But for data analysis, the differences between R and Python are starting to break down.
  - https://www.guru99.com/r-vs-python.html

RStudio


Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
R SCRIPT
Environment
History
Git
Import Dataset
List
Global Environment
Data Objects
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Documents
R seminar series
2.1
Name
Size
Modified
2.1.Rproj
205 B
Aug 26, 2016, 12:52 PM
File ,Plots, Package installation and Help panels
Console ~/Documents/R seminar series/2.1/
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'citation()'
Type 'demo(
'help.start
Type 'q()'
During sta
1: Setting
2: Setting
3: Setting LC_TIME failed, using "C"
4: Setting LC_MESSAGES failed, using "C"
CONSOLE (Script output panel)

Crisp DM framework



Cross Industry Standard Process: CRISP-DM (*cont'd*)

(1) Business/Research Understanding Phase

- Define project requirements and objectives
- Translate objectives into data mining problem definition
- Prepare preliminary strategy to meet objectives

(2) Data Understanding Phase

- Collect data
- Assess data quality
- Perform exploratory data analysis (EDA)

(3) Data Preparation Phase

- Prepares for modeling in subsequent phases
- Select cases and variables appropriate for analysis
- Cleanse and prepare data so it is ready for modeling tools
- Perform transformation of certain variables, if needed

Cross Industry Standard Process: CRISP-DM (*cont'd*)

- (4) Modeling Phase
  - Select and apply one or more modeling techniques
  - Calibrate model settings to optimize results
  - If necessary, additional data preparation may be required for supporting a particular technique
- (5) Evaluation Phase
  - Evaluate one or more models for effectiveness
  - Determine whether defined objectives achieved
  - Establish whether some important facet of the problem has not been sufficiently accounted for
  - Make decision regarding data mining results before deploying to field

Cross Industry Standard Process: CRISP-DM (*cont'd*)

- (6) Deployment Phase
  - Make use of models created
  - Simple deployment example: generate report
  - Complex deployment example: implement parallel data mining effort in another department
  - In businesses, customer often carries out deployment based on your model

Reading:

USING DATA MINING FOR BANK DIRECT MARKETING:
AN APPLICATION OF THE CRISP-DM METHODOLOGY