AD 571: Business Analytics Foundations

LCR Session (Entire Class) Week 3 (05/23/2023)

AGENDA

1 Summary Weeks 1 & 2: Q & A

2 Preparation for Week 3:

A Introduction to Lecture 05 and Lecture 06

B **Individual Exercise: (1)** Tutorial for AD571: Preparation for Assignment 3 — Descriptive Analytics
(2) M3 Module Exercise (Due: 5/29/2023)

C Group Discussion Forum W3, M3 & Quiz 3: Q & A

3 Assignment 3 (due 05/29/2023, 11:59 PM ET): Q & A

Summary Weeks 1 & 2: Q & A

Syllabus: Questions?

Modules 1-2: Questions?

Tutorials:

V-Labs: Technical Solutions

Group Discussion Forum W1 & W2
> **D1: 2.94/4 (73.60%)**
> **D2: Graded by Friday (5/26)**

Quizzes: **Q1: 3.63/4 (90.87%)**
> **Q2: 3.39/4 (84.76%)**

Assignment 1 & 2: Grades Pending

Required Tutorials

Build Content ⌄   Assessments ⌄   Tools ⌄   Partner Content ⌄

Tutorial for AD571: Business Running Case: New York City Real Estate

Tutorial for AD571: Assignment 1

Tutorial for AD571: Assignment 2

**Tutorial for AD571: Assignment 3**

Tutorial for AD571: Assignment 4

Tutorial for AD571: Assignment 5

Preparation for Week 3:

Module 3

May 23 – May 29

Topics:	Lecture 5: Descriptive Business Analytics: Basic Concepts and Applications
	Lecture 6: Modeling and Applications of Probability Distributions
Readings:	Lecture 5:
	• Lecture Notes
	• Evans, Chapter 4
	Lecture 6:
	• Lecture Notes
	• Evans, Chapter 5
Tutorials:	Assignment 3 Tutorial: Descriptive Analytics
	K-Means Clustering
	T-Test Analysis in R
Discussions:	Discussion 3
	• Initial post due Thursday, May 25 at 11:59 PM ET.
	• Respond to at least two of your classmates' posts by Monday, May 29 at 11:59 PM ET.
Assignments:	Assignment 3: Descriptive Analytics, due Monday, May 29 at 11:59 PM ET
	Evans Textbook: Module 3 Exercises, due Monday, May 29 at 11:59 PM ET
Assessments:	Quiz 3, available from Saturday, May 27 at 9:00 AM ET to Monday, May 29 at 11:59 PM ET
Live Classroom:	Module 3 Live Classroom Session — Tuesday, May 23, at 8:00 PM ET
	Technical Session — Thursday, May 25, at 8:00 PM ET

Figure 5.1: Connecting Managerial Decision-Making Process with Business Analytics Process and the Implementation of Selected Business Analytics Methods & Models: Descriptive Analytics
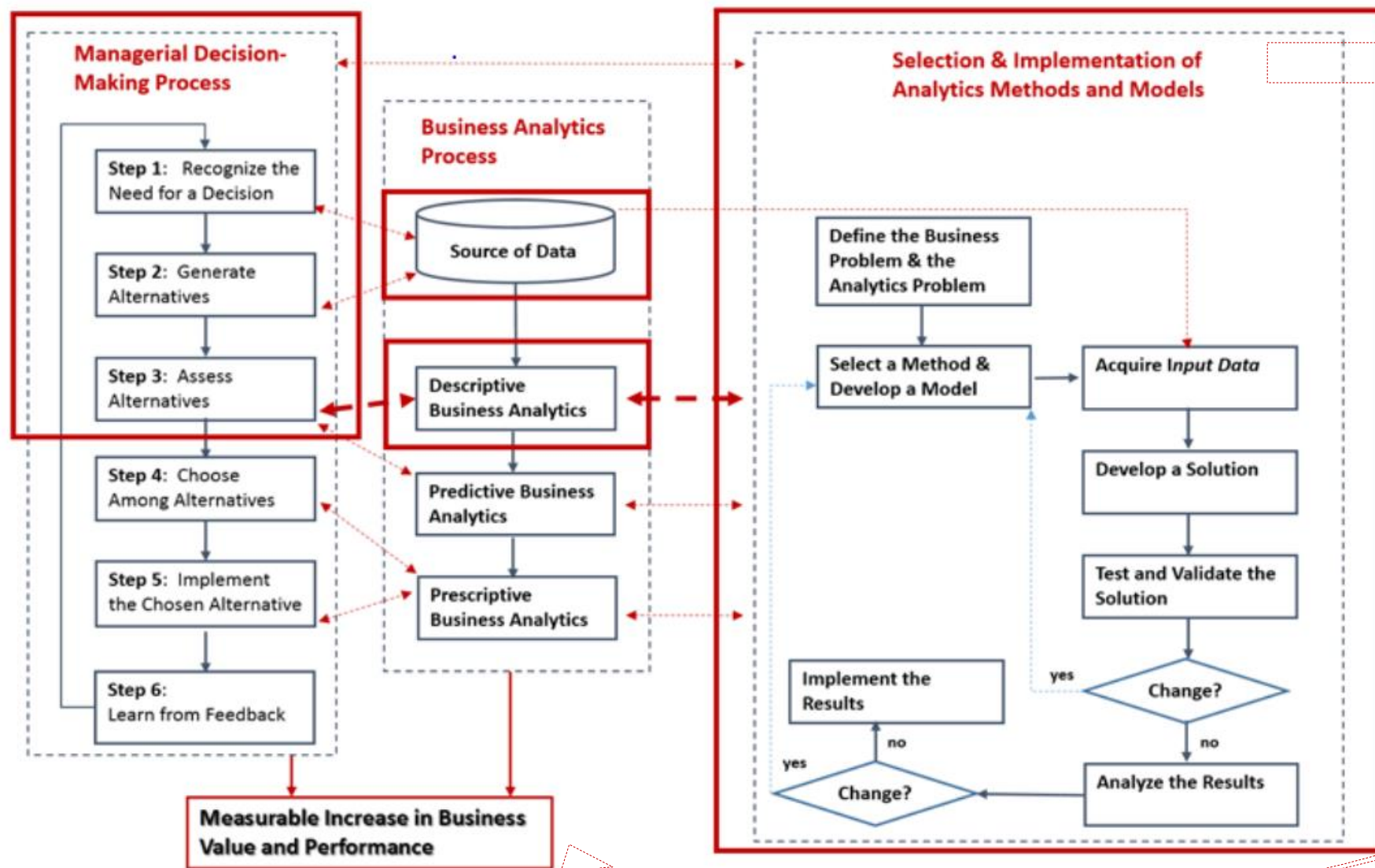
Descriptive Statistics

Clustering

Segmentation

Profiling

t-Test Analysis

Evaluation of the Results: Working with KPI's

Recommendations:
→ Business Analytics Process (overall & next steps)
→ Managerial Decision-Making Process (overall & next steps)

Simplest form of descriptive analytics would be a **historical perspective** on company performance such as sales, turnover, churn, finances, etc.

Goals:
Analysis of historical data
Identification of patterns
Identification of trends
Identification of structure in the data
KPI analysis
Frequency analysis
Root cause of failure

Answers to Questions:
How much was sold in a market or region?
What was the revenue last Quarter/Year?
Where is the lowest productivity?
What are our customer segments?
Why things are happening?
Who are my best customers?
What is the difference between customers?

Preparation for Week 3: Descriptive Analytics

Standard Tools

Reports
Dashboards
Drill-Down
Visualization
Business Intelligence Platform
Spreadsheets
Statistical Software

Statistical Methods

Variation
Central Tendency
Distributions
Correlations
Exploratory Factor Analysis
Principal Component Analysis
Trend Analysis
Cluster Analysis

Descriptive Statistics
Normality Analysis
Homogeneity Analysis
Statistical Inference
Hierarchical Cluster Analysis
Sentiment Analysis
Segmentation Analysis
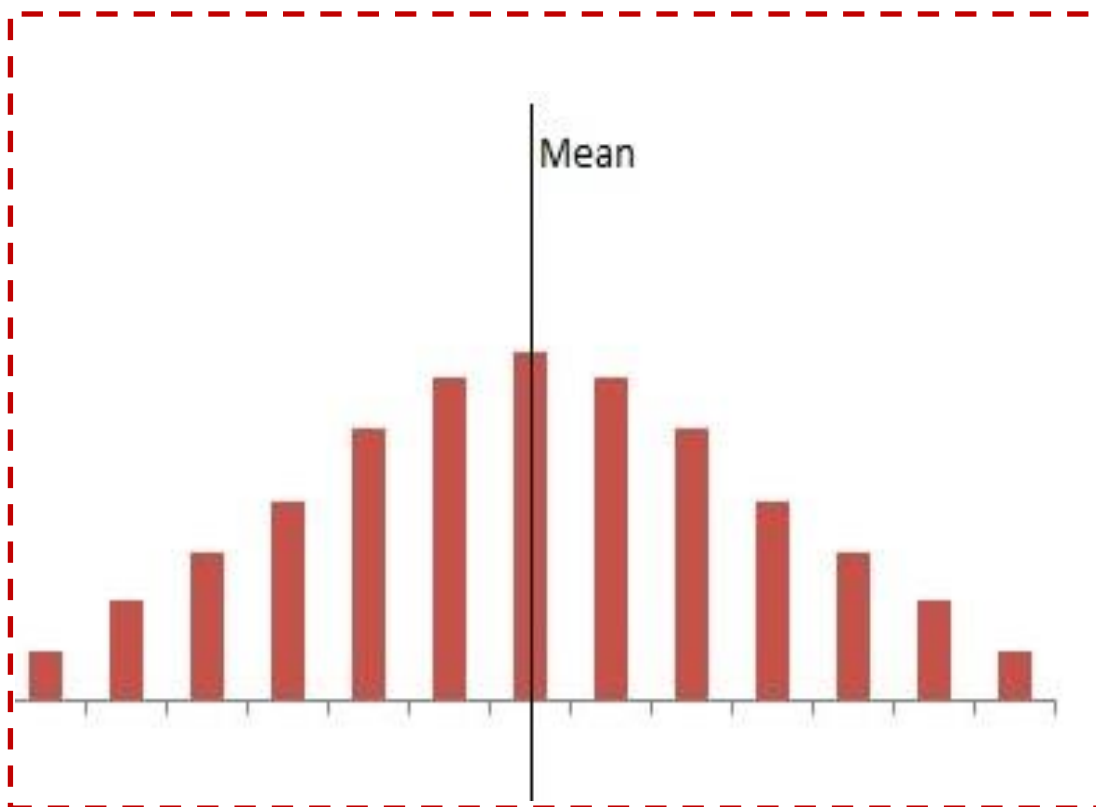
4 C's of Descriptive Analytics:

1. Category (Categorize)
2. Characteristic (Characterize)
3. Consolidation (Consolidate)
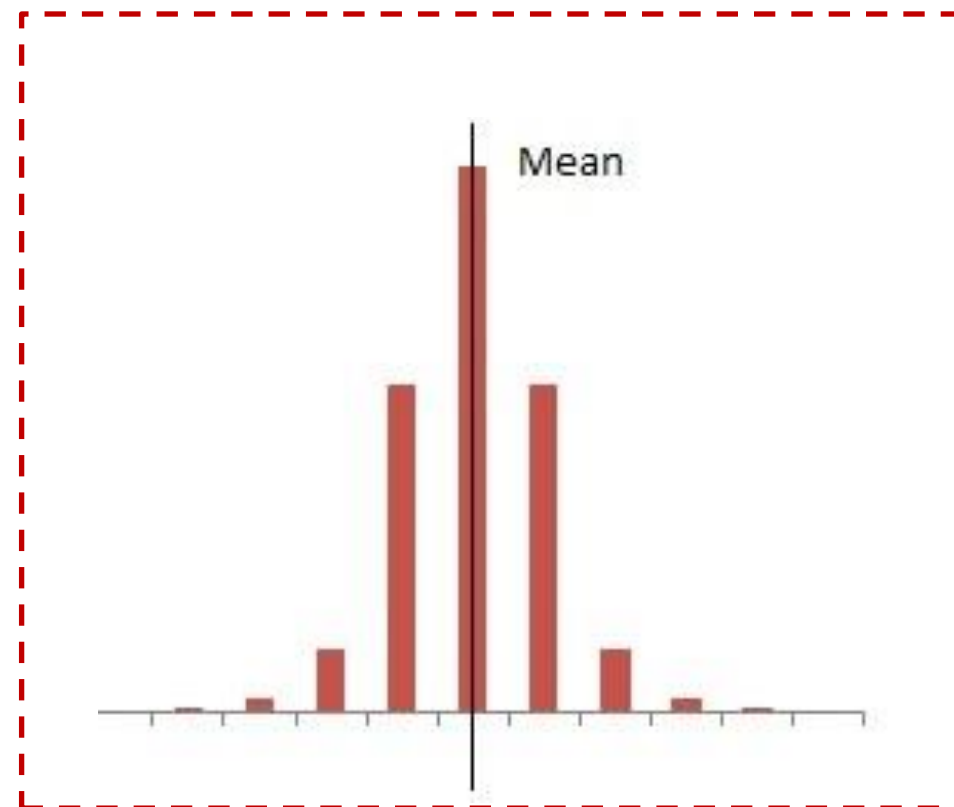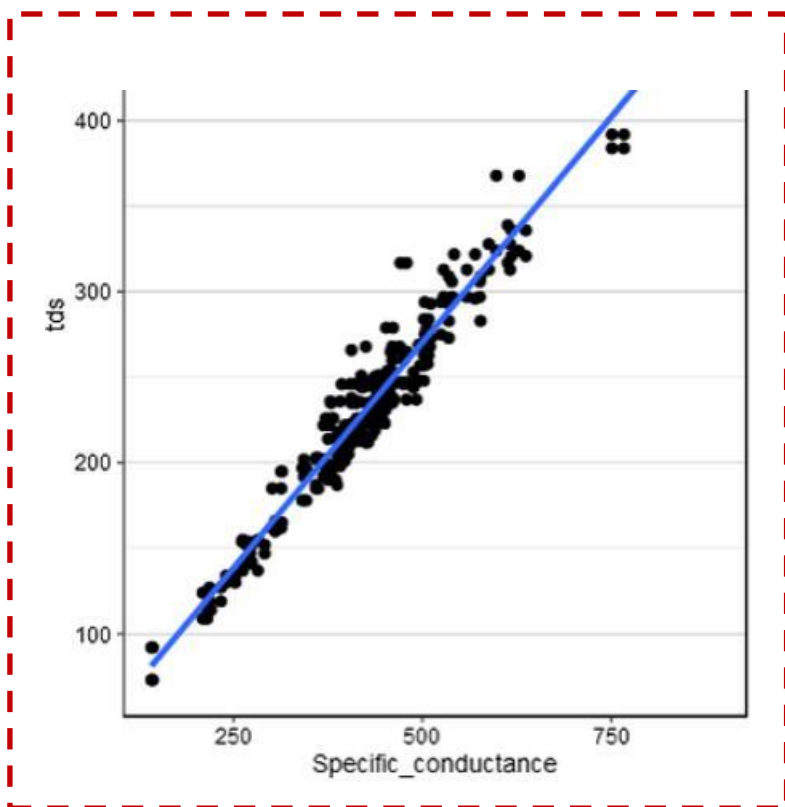4. Classification (Classify)

Histogram of observed data

Source: https://cran.r-project.org
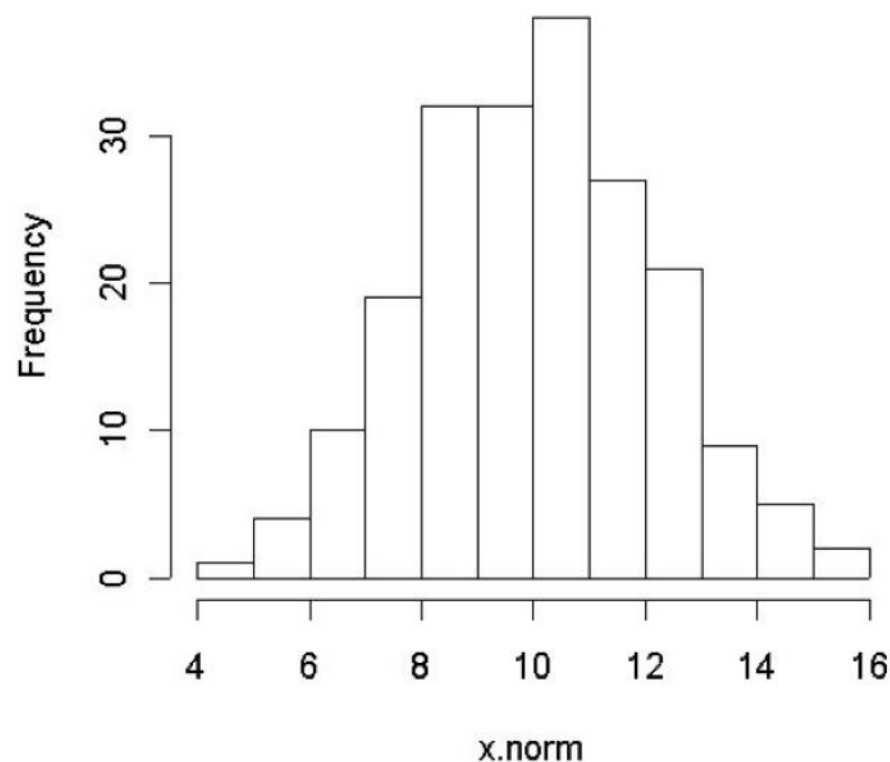
Large Standard Deviation

Small Standard Deviation

BU | Metropolitan College

Preparation for Week 3: Descriptive Analytics: Correlation

Strong Positive Correlation

Source: https://community.rstudio.com

Strong Negative Correlation

Source: https://community.rstudio.com

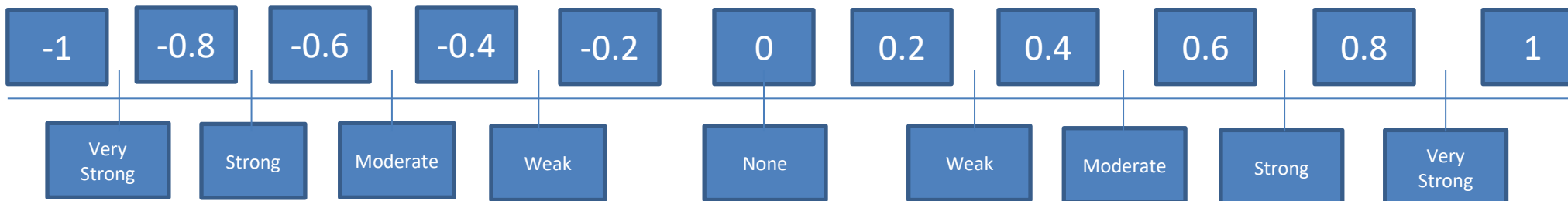| -1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	1
Very Strong	Strong	Moderate	Weak		None		Weak	Moderate	Strong	Very Strong
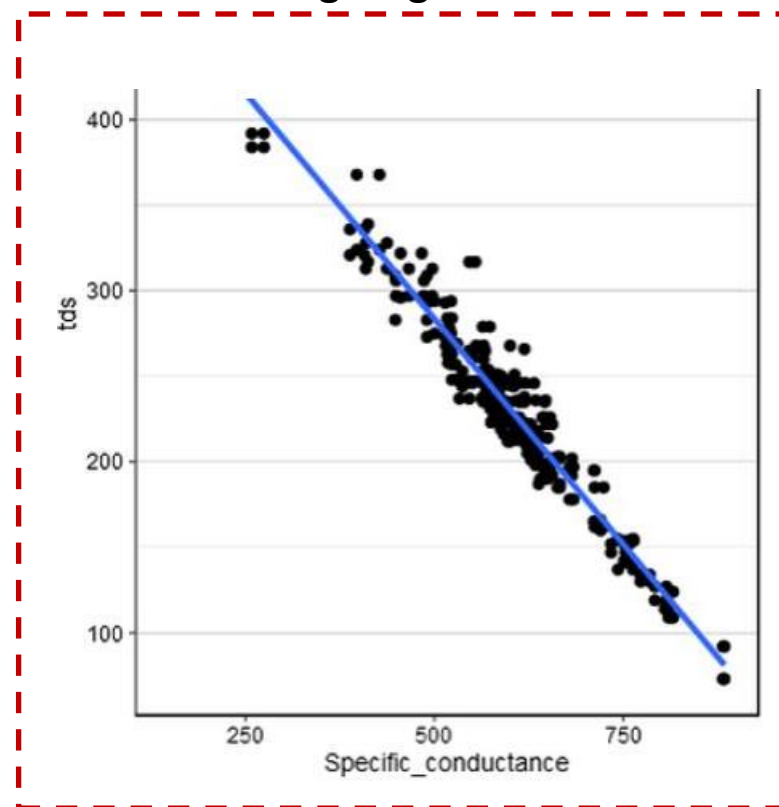
Although descriptive analytics is important to development of answers to some questions, the value derived is **in the way descriptive analytics serves as a precursor to predictive analytics**.
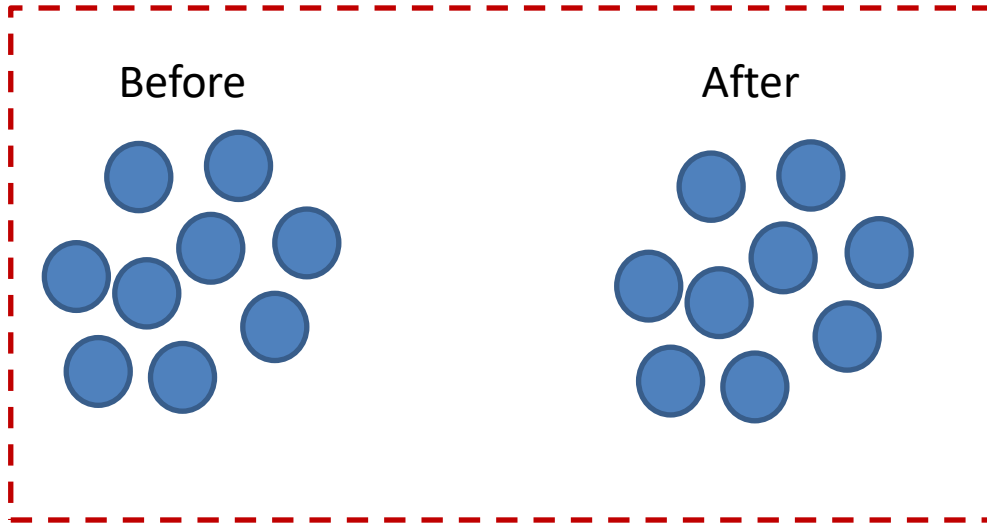
The goal is to leverage past information to forecast possible future outcomes.

## Benefitting from analytics: Increased asset utilization

The Dow Chemical Company sought to increase facility utilization across its office and lab space while improving capital planning, real estate lease management, operations, facility maintenance and energy consumption on a global basis. Using descriptive analytics, Dow was able to produce reliable data for decisions that helped the company identify under-utilized space, achieve a 20 percent increase in facility use and generate a savings of approximately USD4 million annually via space consolidation.

By examining key metrics and key performance indicators of energy use or vehicle maintenance, for example, descriptive analytics can produce indicators such as cost per square foot, kilowatt hours per person or mean time between failures for specific asset issues. By combining information from different, often disconnected sources and then comparing and contrasting data, descriptive analytics can provide a comprehensive view and context for what has happened, as well as current asset status.

Source: IBM Watson IoT (*https://www.ibm.com/downloads/cas/3V9AA9Y5)*

Preparation for Week 3: t-Test

Before   After

Group 1   Group 2

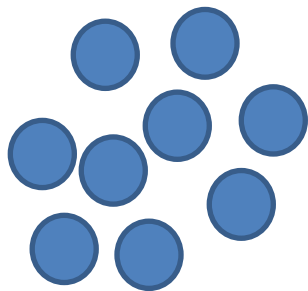Based on sampling, is there a statistically significant difference between a group before and after exposure to a manipulation?
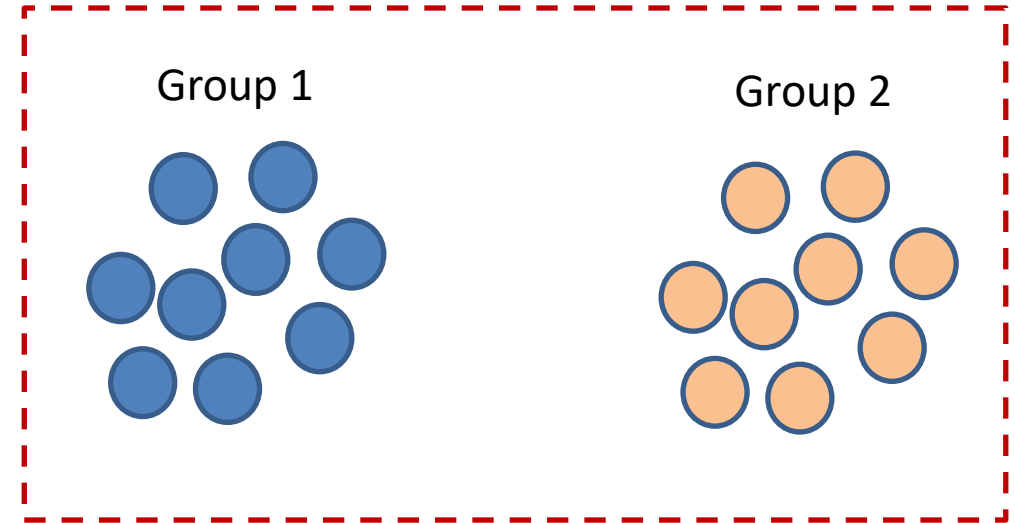
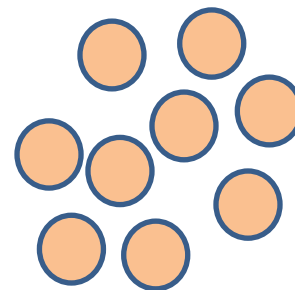Based on sampling, is there a statistically significant difference between 2 groups?

Think of a few examples of interest to you…

Mean of Neighborhood X   Mean of Neighborhood Y

```
# T-Test with NEighborhood X;
# H0 = Both are equal, or there is no difference
# Ha = NEighborhood X less than the mean price of Neighborhood Y,

NeighborhoodX<-filter(.data = NYCHistorical,SalePrice!=0,NbhoodID==999)
NEighborhoodY<-filter(.data = NYCHistorical,SalePrice!=0,NbhoodID==000)
t.test(x=NeighborhoodX$SalePrice,y=NeighborhoodY$SalePrice,alternative = "less",mu=0,conf.level = 0.95)
```
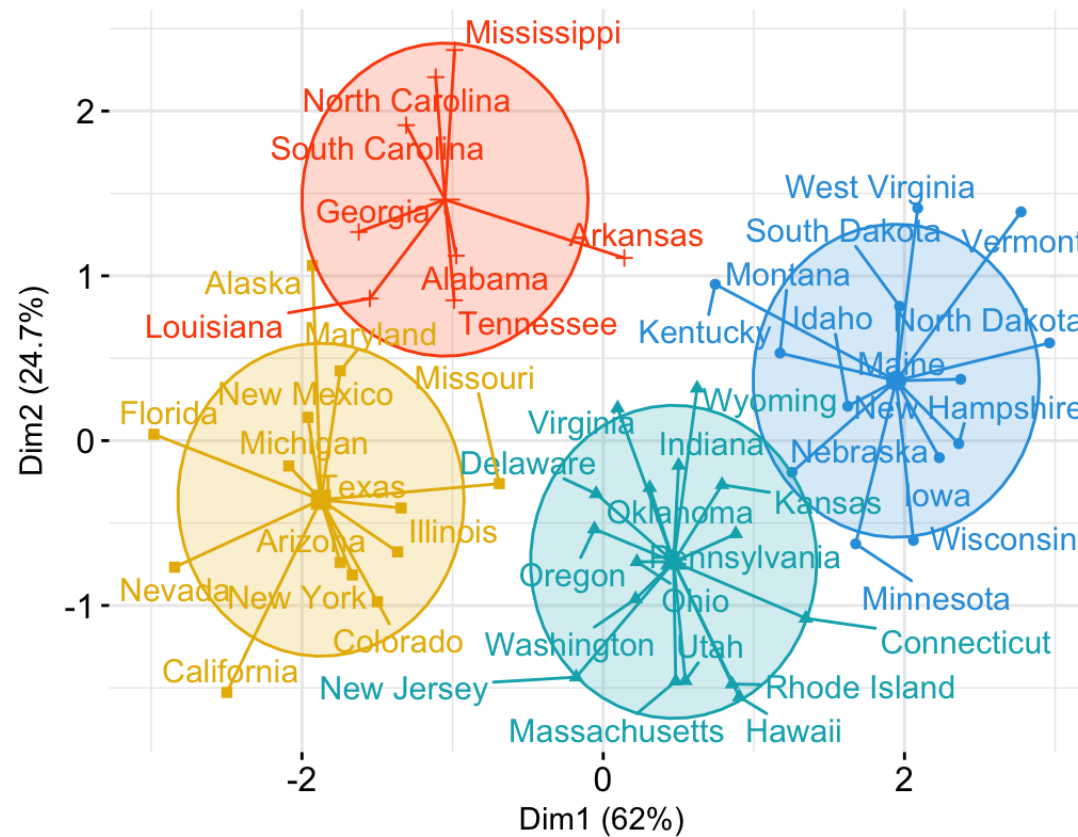
```
t = -9.3172, df = 2593.8, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -392413.4
sample estimates:
mean of x mean of y
 800816.2 1277394.5
```

DEMO

BU | Metropolitan College

Source: https://stats.stackexchange.com
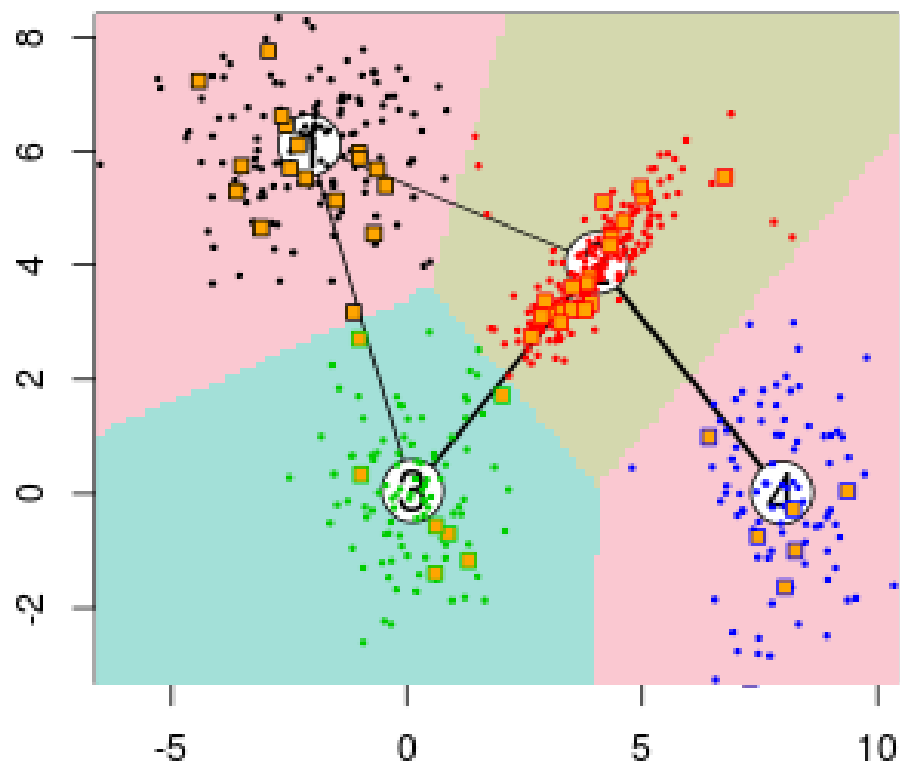
Partitioning Clustering Plot

Source: http://R-blogger.com

cluster [●] 1 [▲] 2 [■] 3 [╫] 4

Identification of communities in social networks

http://hastac.org
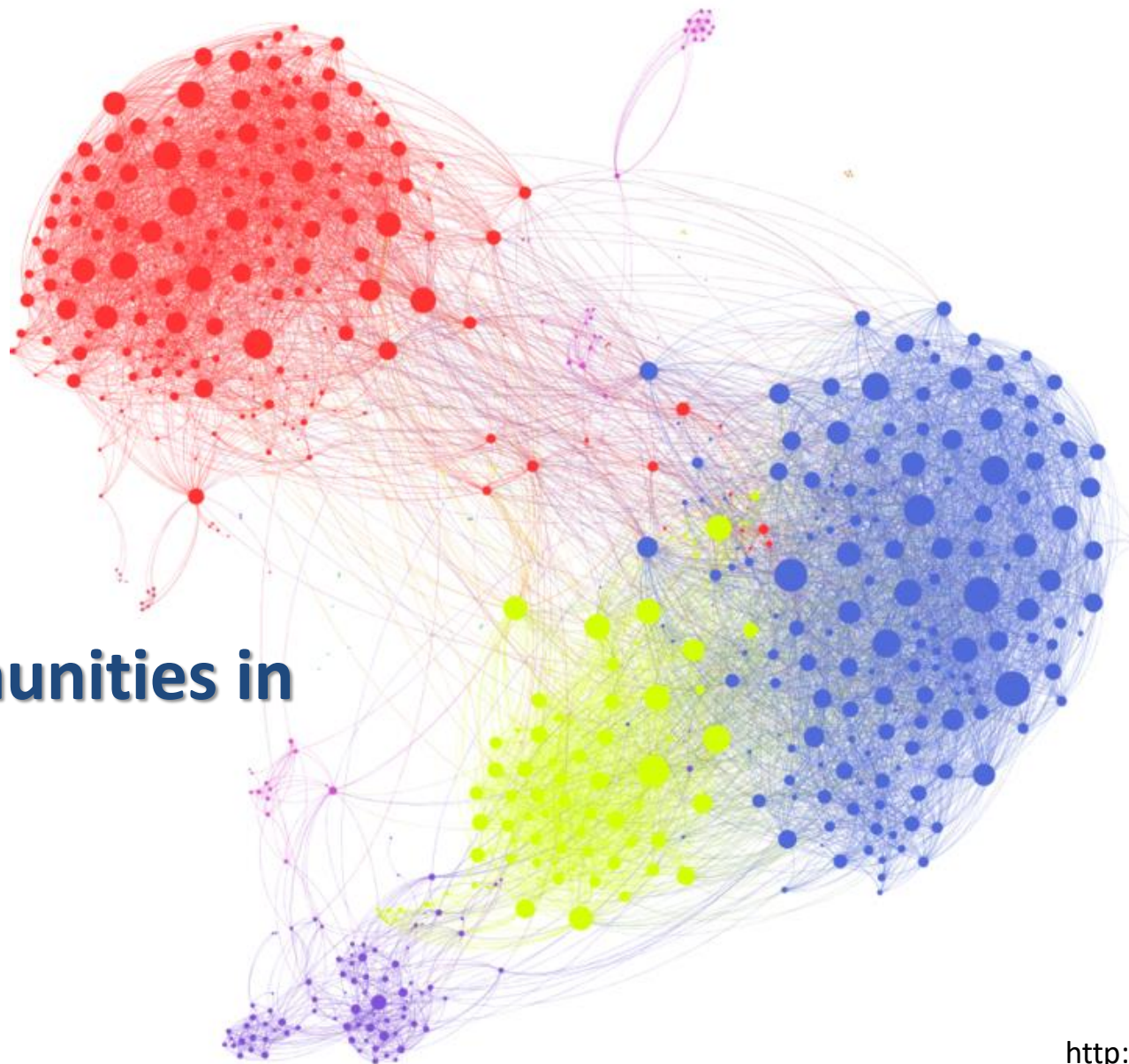
Google Search Results

Finds meaningful and relevant groups and collections or results

http://searchenginemarketingland.com
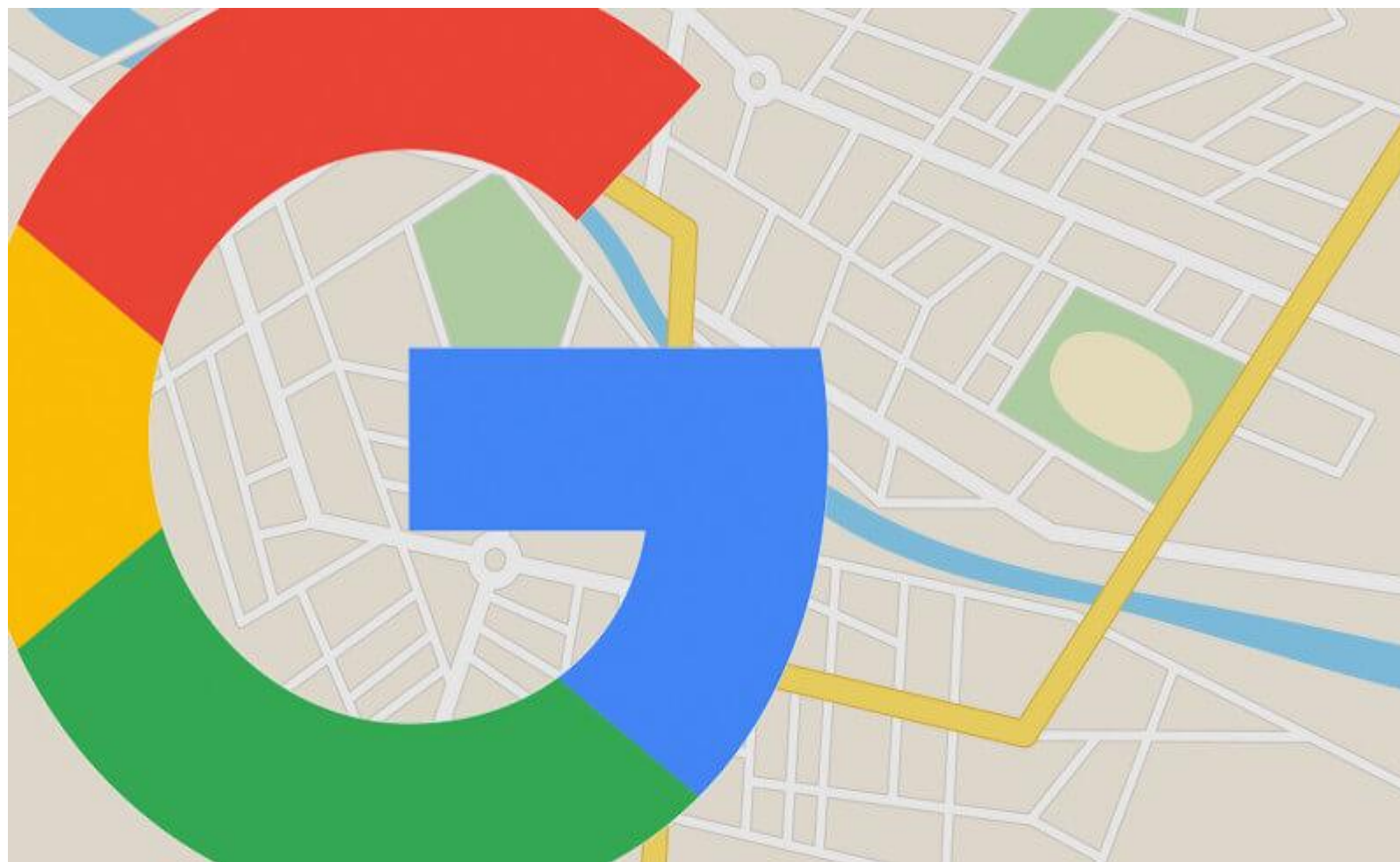
Market research professionals generating clusters from:

- **Survey Data**
- **Panels**
- **Polls**
- **Transaction Data**
- **Behavior Data**
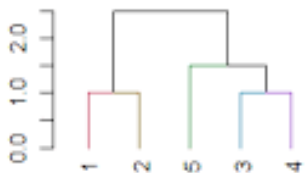- **CRM Data**

	A	B
1	Survey ID	Q6-Organic
2	1	3
3	2	4
4	3	4
5	4	5
6	5	5
7	6	5
8	7	4
9	8	4
10	9	4
11	10	5

http://researchgate.net

Types of Clustering:

1. Hierarchical Clustering – Useful for organization data and the tree can be stopped at any level.
2. Distribution-Based Clustering – Probability of belonging to a cluster decreases with distance from center.
3. Density Based Clustering – High density of points are placed into clusters. Useful for low dimensional data
4. Centroid Based Clustering – High efficiency and simpler clustering process but not the best when there are outliers.

1

Source: https://cran.r-project.org

2

3

Source: https://cran.r-project.org

4

Source: https://cran.r-project.org
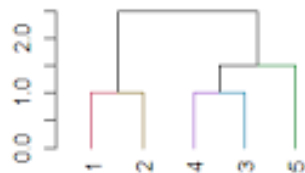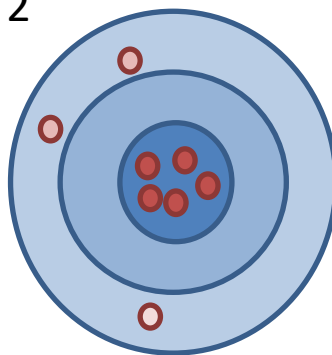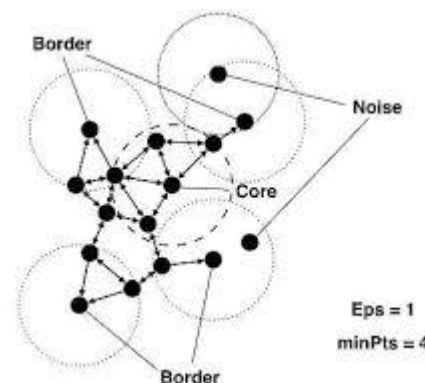
DEMO

```
#K-Means Clustering
DataFrame1<-filter(NYCHistorical,SalePrice!=0,GrossSqFt!=0,TotalUnits!=0)%>%
  as.data.frame()

DataFrame2<-group_by(DataFrame1,NbhoodID)%>%
  summarise(avgSalePrice=mean(SalePrice),avgGrossSqFt=mean(GrossSqFt),TotalUnits=sum(TotalUnits))

zscores<-scale(DataFrame2[c(-1)])

k<-kmeans(zscores,centers=5)
cluster<-cbind(k$cluster,DataFrame2)
view(cluster)

ggplot(cluster)+geom_point(mapping = aes(x=TotalUnits,y=avgSalePrice,size=avgGrossSqFt,color=k$cluster))
```
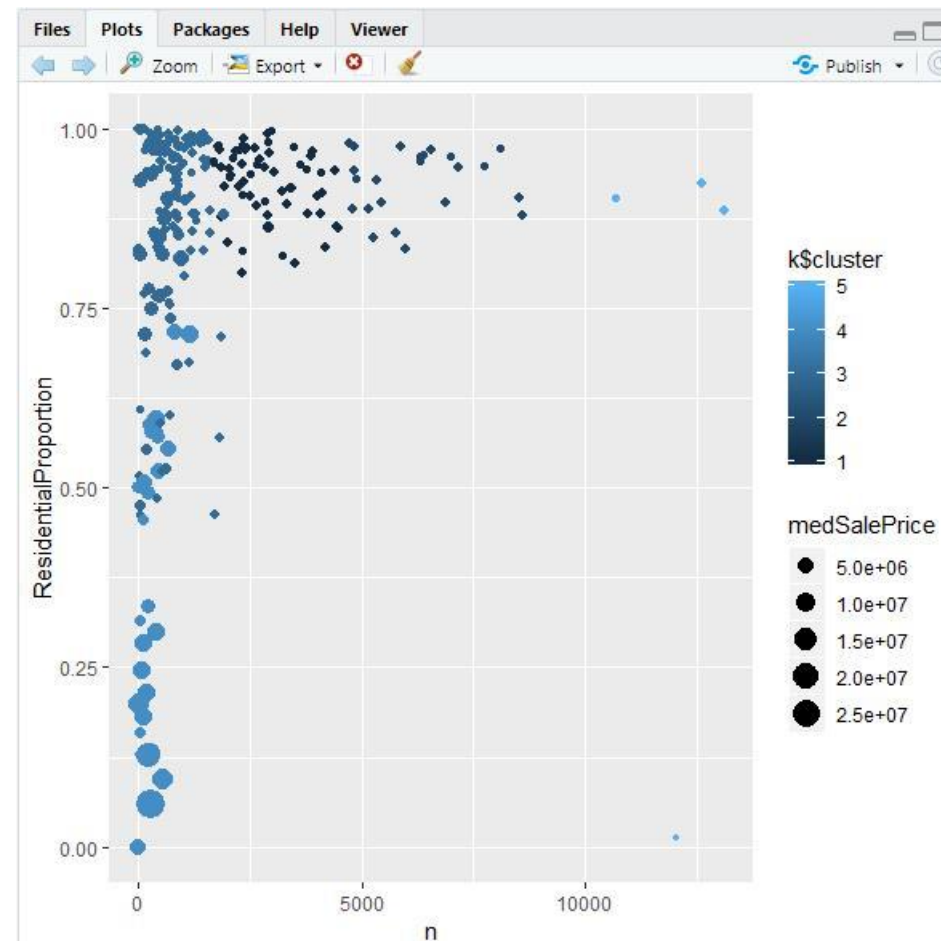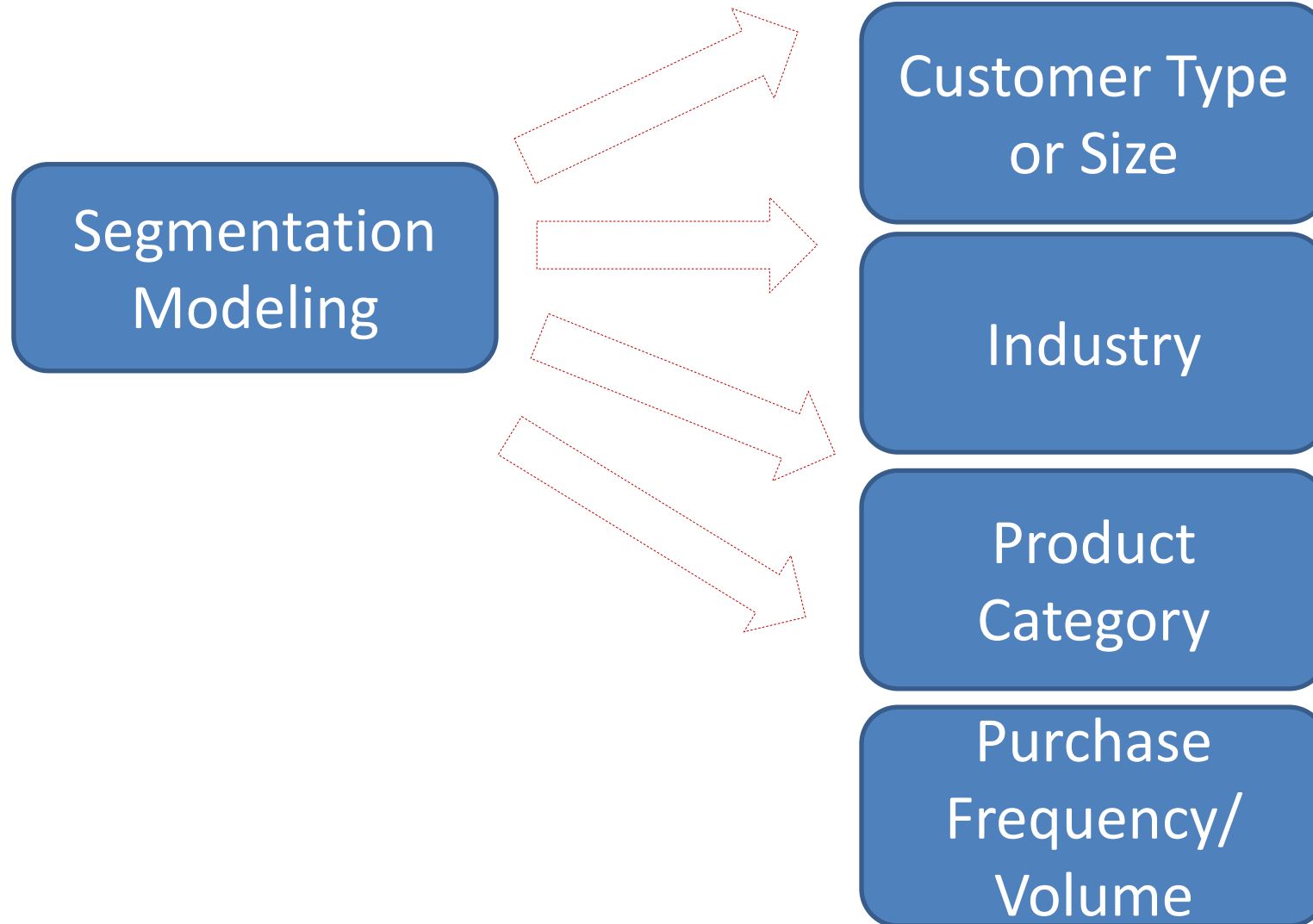
Use Cases of Segmentation:

- **Email Marketing**
- **Digital Advertising**
- **Prospecting**
- **Customer Analytics**
- **Creative Decisions**
- **Communication Decisions**

Behavior

Detailed Revenue Attribution

Detailed Profit Attribution

Detailed Cost Attribution

Specific Products Sold

Preparation for Week 3: Group Discussion Forum W3

Please select just one of the bullet points listed below.

- Discuss the benefits of the different descriptive analytics techniques and how they can be integrated into the decision process
- Discuss the most useful dashboard elements or visualizations in your profession
- Discuss the meaning of Business Intelligence and its relevance to descriptive analytics
- Explain the significance of descriptive analytics to your field of work
- Discuss the possible applications of tools from Lectures 5 and 6
- Discuss topics of your choice from Lecture 5 and Lecture 6

BU Metropolitan College

Quiz 3 consists of 10 multiple choice and true-false questions.

Quiz 3 will cover information from the
➢ Online lecture notes (Lecture 05 and Lecture 06),
➢ Evans Chapter 4 and Chapter 5

The quiz is open-book.

Format
➢ You will have 30 minutes to complete the quiz. There is a clock in the upper right corner of the screen keeping time for the exam.
➢ You can take the quiz only once.
➢ Each question will be delivered one at a time.
➢ You can revisit the questions and change your answers as many times as you want before submitting the exam.

Preparation for Week 3:   Module 3 Exercise: Q & A

Questions are located at the end of Chapter 4
Section: Problems and Exercises

Group 1 - Roman			Group 2 - Vivek	
Due Date	Monday, May 29, 2023		Due Date	Monday, May 29, 2023
Student	**Chapter/Question**		**Student**	**Chapter/Question**
Yurui Chen	Chapter 4 - #46		Audrey Chan	Chapter 4 - #46
Victor Brice Fedjo Yeme	Chapter 4 - #10		Flori-Ann DeLa Cruz	Chapter 4 - #10
Jonathan Garrison	Chapter 4 - #13		Marcel Fernandes Silva	Chapter 4 - #13
Payton Hatcher	Chapter 4 - #40		Avirul Islam	Chapter 4 - #40
Faria Hossain	Chapter 4 - #48		Jiarui Lin	Chapter 4 - #48
Jiamin Li	Chapter 4 - #7		Joyce Machau	Chapter 4 - #7
Kaitlynn Nguyen	Chapter 4 - #44		Nicole Matarazzo	Chapter 4 - #44
Miranda Petrillo	Chapter 4 - #53		Timothy Olakunle	Chapter 4 - #53
Haoqiang Qi	Chapter 4 - #57		Olu Olayeye	Chapter 4 - #57
Nesteshia Riddell-Dell	Chapter 4 - #62		Sandhya Ramani	Chapter 4 - #62
Cassandra Simoneau	Chapter 4 - #51		Sri Amruta Sripada	Chapter 4 - #51
Yang Yang	Chapter 4 - #63		Samuel Stevens	Chapter 4 - #63
Samuka Yekeh	Chapter 4 - #44		Jack Swartz	Chapter 4 - #44

Business Running Case–Targeted Outcomes:

Application of Descriptive Analytics

BOSTON UNIVERSITY

W-2
- **Term Project A1**: Data Import and Visualization
- **Term Project A2**. Data Manipulation In R

W-3
- **Term Project A3**. Descriptive Analytics Techniques In R

W-4
- **Term Project A4**. Predictive Analytics Techniques In R

W-6
- **Term Project A5**. Prescriptive Analytics Techniques in R and Excel

W-7
- **Term Project A6**. **Term Project Final Presentation** Model Deployment

V-Lab Instructions

All course tools and SQL Server can be accessed from the V-Lab, which is included in tuition.

Assignment 3 Requires access to R Studio and SQL Server

Assignment 3: Descriptive Analytics

Assignment 3 Objective: Prepare a managerial report, starting with an executive summary; expected length up to 3-4 pages APA format, excluding cover page, table of content, and appendixes.

1. Provide several descriptive statistics for real estate sales in your neighborhood. Include:
   a. The total number of sales in your neighborhood since 2009
   b. The mean sale price and gross square footage for residential properties in your neighborhood since 2009
   c. A five-number summary for both sale price and gross square footage for residential properties in your neighborhood since 2009
   d. The proportion of sales of residential, commercial, mixed, and other properties in your neighborhood since 2009
   e. The standard deviation of sale prices for residential properties in your neighborhood since 2009
   f. The correlation between sale price and gross square feet for residential properties in your neighborhood since 2009
2. Perform k-means clustering, comparing your neighborhood to other neighborhoods. Choose at least 3 of the following KPI's
   a. Median Sale Price for residential properties since 2009
   b. Number of sales for residential properties since 2009
   c. Standard deviation of sales of residential properties since 2009
   d. Proportion of residential sales since 2009
   e. Price of 1 gross square foot of residential real estate since 2009
3. Choose one other neighborhood, and test the hypothesis that, starting in 2009, the average residential property costs more, less, or a different amount in your neighborhood (Note: this requires a t-test.)
4. Write 3-4 pages summarizing your findings with a focus on the output, interpretation of the output, and what the insights mean for our decision-making process.

LOCATION OF THE INFORMATION AND SOFTWARE

Course Website

Page on Blackboard: "Assignments">>Assignment 3

Tutorials Section

Link
Tutorial – Assignment 3

Q & A