

# MIS 0855 – Data Science (Section 006) – Fall 2017

## Exam #3 Guide

Read class slides, reading assignments, and in-class exercises. Take a look at videos posted on the class site.

### Day 1 – Introduction

- The meaning of “connecting the dots” in the NYPD video
- Why we need information and data for?

### Day 2 – Data and Science

- The three requirements of a hypothesis
- Examples of rationales for hypotheses

### Day 3 – Data, Information, and Knowledge

- The difference in data, information, and knowledge
- Examples in data, information, and knowledge
- Figures in slide p. 10

### Day 4 – Data and Metadata

- The meaning of metadata or data dictionary
- Examines of metadata
- Why we need metadata

### Day 10 – In Data We (Mis)Trust

- The problems with UCR Data
- Why the unemployment rate data could be misleading
- Why reviews and ratings at Yelp or Amazon.com could be untrustworthy
- The ways data becomes untrustworthy

### Day 13 – Data Visualization

- Four data visualization principles
- Key messages from visualization in Slide Page 5-10
- When to use a table instead of a graph. When to use a graph instead of a table
- When to use a bar chart, a line chart, a pie chart, and a scatter plot, for what?

### Day 14 – Data Distortion

- Watch the YouTube video for Day 14
- A pie chart with the total percentages greater than 100%
- A bar chart without 0 in the vertical axis
- The correct size of squares
- State leader vs. national leader

- Why a 3-D chart is misleading
- Why using as little link as possible

#### **Day 15 – In-Class Exercise – Creating Interactive Dashboards**

- What a Tableau extract is
- What Tableau Public is for

#### **Day 16 – In-Class Exercise – Integrating Data**

- The calculated field for RepAge
- How to display a trendline in a graph

#### **Day 17 – Dirty Data**

- Watch the two videos for Day 17
- Why data gets dirty
- The origin of dirty data
- The agency problem and a solution to fix that problem
- What to do with outliers (usually large data)
- Activities in data cleansing
- Five characteristics of high-quality data
- Trade-off in data cleansing

#### **Day 18 - In-Class Exercise – Finding Bad Data and Assignment #3**

- How to use MATCH function in Excel
- Using a scatterplot to find outliers

#### **Day 21 - In-Class Exercise – Visualizing KPIs**

- How to visualize the departure delay KPI in Tableau
- How to create a heatmap in Tableau

#### **Day 22-23 - In-Class Exercise – Visualizing Maps**

- What types of maps are available in Tableau – state, city, county, metropolitan areas, zipcode, congressional districts, etc.
- Difference between a filled map and a symbol map
- How to fix missing (or “unknown”) geographic locations

#### **Day 25-26 – In-Class Exercise – Advanced Data Visualization**

- How to create a dual line chart
- What information is available from a box plot
- What is the name of this chart? (among dual line chart, bubble chart, box plot, bullet graph, histogram)

#### **Day 27-28 – In-Class Exercise – Visualizing Network**

- How to structure an Excel file for visualizing a network in Tableau
- Steps in visualizing a network in Tableau (Part 1)

**Day 31 – Predictive Analytics**

- Examples of predictive analytics
- Collaborative filtering
- Measuring customer value with R.F.M.
- Three types of predictive analytics
- Association detection
- Regression
- Decision tree

**In-Class Exercise – Day 32 – Simple Predictive Analytics**

- Prediction interval (or confident interval)
- How to conduct an association analysis (association detection) in Tableau
- Outer join – how we joined VadelayOrders records for association detection (Part 2)

**Day 35-36 – Data Science and Your Career**

- The causes of the great decoupling
- Beveridge curve
- The impact of technology development on employment