

Final Project



Purpose

* To put to work the tools and knowledge that you gain throughout this course.
* Benefits:
  + More experience with data
  + Self directed learner
  + Working with others
  + Building data science portfolio



Project Goal

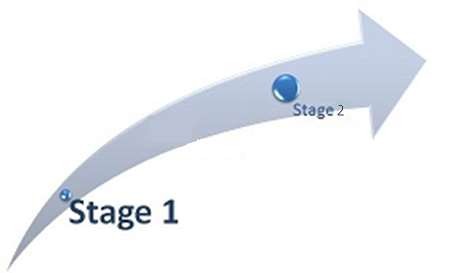
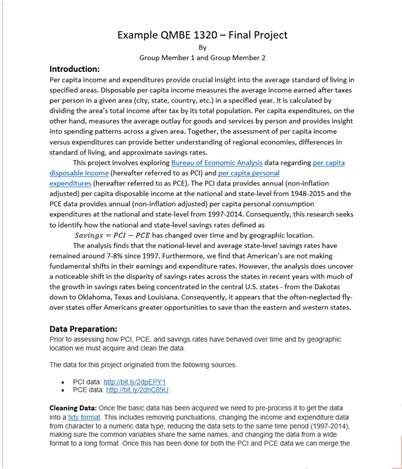
* The principal goal of this project is to:
  + Import a real-life data set,
  + Clean and tidy the data
  + Perform basic exploratory data analysis

All while using Excel and Word to create a report that is clean and professional



Project Data

* Select 1 Data Set
  + Dog Data (Petfinder.com)
  + Hotel Data (Bookings.com)
  + NFL Data
  + Spotify Data
* NOTE:
  + There are a lot of variables!
  + Do not get overwhelmed!
  + Look them over and select which ones you want/need
  + Think about how you can use some to create new variables



What is it?

* A Report! - An analysis on real-life data
  + Conducted **individually** or with a

**partner**

* Expectations
  + Neat
  + Well-written (Grammar, punctuation etc.)
  + Professional
  + Tables and Graphs meet class tips and expectations



Stages (“Exam 3” and “Final Exam”)

* Stage1(Exam 3):
  + Why you chose certain data
  + What problem statement do you want to solve or investigate?
  + Clean Data
  + Visualize Data
* Stage 2(Exam 4):
  + Stage 1 + Data Analysis and Summary
* Full outline and rubrics are posted for the assignment on Canvas



You Decide What You Want to Study!

* Problem Statements:
  + 1. Identify the Problem
    - Specific
    - Relevant
    - Actionable
  + Bad Example:
    - I want to analyze customer satisfaction
  + Good Example:
    - I want to find out how customer satisfaction affects retention and revenue for our online store.
    - I want to increase sales by identifying the factors that influence customer behavior.
    - I want to understand the relationship between air pollution and respiratory diseases.



You Decide What You Want to Study!

* Problem Statements:
  + 2. State the Objective (SMART)
    - Specific
    - Measurable
    - Achievable
    - Relevant
    - Time-Bound
  + What do you want to achieve or learn from this analysis?
  + Good Example:
    - I want to determine how customer satisfaction influences retention and revenue, and identify the factors that affect customer satisfaction



You Decide What You Want to Study!

* Problem Statements:
  + 3. Define the Scope
    - Narrow your:
      * Sources
      * Time Period
      * Target Population
      * Variables
      * Assumptions
  + Good Example:
    - I will use survey data from the past six months, segment customers by purchase frequency and amount, and assume that satisfaction is measured by the Net Promoter Score
    - I will use Minnesota state respiratory and pollution data from the past six months to study the relationship between air pollution and disease.



You Decide What You Want to Study!

* Problem Statements:
  + 4. Formulate the Question
    - Convert your objective to a specific, measurable, and testable question.
  + Good Example:
    - How does customer satisfaction impact retention and revenue for our online store, and what are the main drivers of customer satisfaction?



You Decide What You Want to Study!

* Problem Statements:
  + 5. Choose data and Tools
    - Which variables are you going to use?
      * Can you or do you need to create more variables
    - Good Example:
      * Increase sales: use sales data, customer data, marketing data
    - Which methods are you going to use?
      * Correlation, regression, time series, prediction and forecasting?
    - Good Example:
      * Air pollution and disease: correlation plots, regression etc.



You Decide What You Want to Study!

* Problem Statements:
  + 6. Review and Refine
    - Check your problem statement for:
      * Clarity
      * Conciseness
      * Completeness
    - Make necessary adjustments or improvements
  + Helpful Hints
    - You could ask others for feedback
    - Test your assumptions
    - Compare



Other Helpful Things

* Create a Story!
  + Logical, cohesive (Not just a bunch of graphs) – Connect the dots!
  + Data -> Insights -> Actions
* Spend time and effort studying descriptive statistics and visualization
  + Understand the data to help give you insights and the path forward
* Questions to ask your self
  + Do sub groups matter?
  + Why are data missing?
  + Are there trends, patters, seasonality, other noticeable things



Today

* 1. Find a partner or decide to work individually!
* 2. Go through the data files and determine what data you think you want to analyze
* 3. Sign-up at the front
  + You and your partner’s name and which data set you are planning to use
* 4. Begin the analysis
  + Based on the variables you have – what are some problem statements/questions you can answer (2-3)