Project Information

Project Context

Role:
- Your team should act as consultants
- You need to analyze hotel cancelations for your client

Goal:
- How might your client reduce cancelations
- How might your client better predict cancelations
- Others???
- No other specific questions / goals will be provided

What to analyze – should be:
- A function of what the team determines might be useful
- Determined by each project team

Remember this needs to be data driven –
→focus on what data available

Data Info

- Hotel Cancelation Data
- The metadata for this data file can be fond in Blackboard
- Note that some values in the dataset are blank / NA

Kanban Project Methodology

Kanban Overview

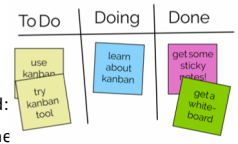
- Originally defined / used in manufacturing
- Gaining traction in software development and data science
- Kanban's Two key Principles:
  1. Visualize the flow
  2. Minimize work-in-progress

By limiting tasks that are being completed simultaneously, Kanban enables agility and prevents overloading the development process

→ *Can use www.trello.com to create a Kanban board, which can be shared*

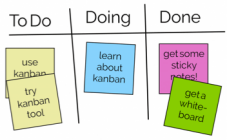
Visualize the Flow

- The team starts with a list of potential tasks in the "To Do" column
- In a simple three column Kanban board:
  - When the team starts working on the task, the Kanban card (task) is moved from the "To Do" to "Doing" column
  - When the team completes its task, it is moved to the "Done" column
- Teams often define more columns (e.g., "validate" before "Done")



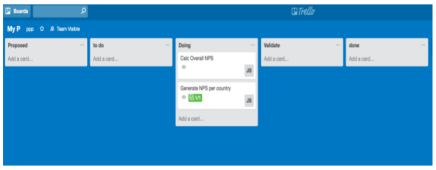
Limit Work in Progress

- Uncompleted work is known as work in progress (WIP)
- WIP limits define the maximum number of tasks that can simultaneously exist in a given column.
- The concept minimizing WIP enables agility since new knowledge is gained prior to the start of more work



Kanban Tasks

- Split the work into pieces (each piece is a task)
- Each team will define what to investigate (ex. “link weather data to our previously collected data”)
- These will be listed (in a prioritized order) in a “to do” column.
- Then, as space permits (based on the # of allowed tasks at each step), the task flows to the next column.



Expectations

1) Work at a consistent pace throughout the semester
2) Tasks should be distributed equally across the team members
3) Tasks should typically not take a long time to complete – one week target, two weeks is fine, but not a month
4) For each task, try to define / explain:
    a) What will be created (ex. a visualization, a predictive model)
    b) What will be observed (ex. data output from a predictive model)
    c) What will be analyzed (ex. is the model predictive, what attributes are important)
5) Tasks should be at an appropriate level of effort / detail
6) There a good list of “to do” items (from the ‘proposed’ column)

Project Grading

School of Information Studies
SYRACUSE UNIVERSITY

Project Deliverables Review

Word Document:

- Target audience is your manager / instructor
  (hint: your manager/instructor is a data science expert)
- Focus on what was accomplished
- Should describe all analysis done, even if an analysis did not generate any interesting results, it should still be included

Presentation:

- Target audience is your client *(hint: the client is not a data science expert)*
- Presentation length is 10 minutes *(lab instructor will explain specifics)*
- Be sure to include the following in your presentation:
  - Number of records in dataset evaluated
  - Key drivers identified; accuracy of results

Timeline

- This week, you will get:
  - The case description
  - Information on the data available
  - Think about :
    - What might be interesting questions, how you might work on the project, what data might be interesting to explore
- Next week, via your lab instructor, you will get:
  - dataset
  - Details on team, etc.

Some Project Hints

- Is your analysis based on absolute numbers (which can be an issue when comparing two populations)
- Think about :
  - When to transform attributes (columns) from numbers to categories ex. low (1), med (2-4), high (5)
  - How "good" are the models (covered later in the semester)
- You need actionable insights!
- Think about what to do for rows that are empty

Presentation (5%)

0.5% - Business Questions - are they appropriate within the context

0.5% - Use of Descriptive statistics - Did you provide context and a basic understanding of the data

1% - Use of modeling techniques - Did you try at least 3 different models and explain why they were/were not useful

1% - Visualization - Did you convey the results in an easy to understand manner

1% - Interpretation of the results/Actionable Insights - Are the results actionable (as compared to just interesting)

1% - Know your audience - did the presentation present findings in an easy to understand way (ex. no data science lingo, easy for others to follow the logic)

Word Document (15%)

1% - Business Questions - appropriate within the context?

1% - Data cleanse/munge/preparation - transform/clean/munge the data appropriately? What about NAs?

1% - Use of Descriptive statistics - provide context and a basic understanding of the data?

4% - Use of modeling techniques – try 3 different models (for possibly different questions, evaluating results correctly)

3% - Visualization- convey results in an easy to understand manner?

4% - Interpretation of the results/Actionable Insights - Are the results actionable (as compared to just interesting)

1% - Validation - How do you know your results were correct (i.e., no errors)

Final Project (Word doc)

Example Table of Contents:

- Introduction (scope/context/background)
- Business Questions addressed
- Data Acquisition, Cleansing, Transformation, Munging
- Descriptive statistics & Visualizations
- Use of modeling techniques & Visualizations
  (noting techniques explored but not used in presentation)
- Actionable Insights / Overall interpretation of results
- Appendix – Code (can be link to the code)