



# **799:470/641 Supply Chain BI/AI Team Project**

- Analytics translators:
  - Define business problems that analytics can help solve
  - Guide technical teams in the creation of analytics-driven solutions  
(knowledge of common programming languages, such as Python and R)
  - Embed solutions into business operations
- How to train analytics translators:
  - Follow an end-to-end process that is applicable to a wide range of business problems

## How to Use the Data Science Workflow Canvas

### Step 1: Identify your problem statement

What problem are you trying to solve? And what larger issues do that problem address? This section helps you address the “why” of your project.

### Step 2: State your intended outcomes/predictions

Yes, you won't know what your outcomes are until after you're done with your project, but you should at least have an idea of what you think they should look like. Identify potential predictor (X) and/or target (y) variables.

### Step 3: Determine your data sources

Where are you sourcing your data from? Is there enough data? And can you actually work with it? Sometimes you might have access to ready-made [datasets](#), or you might need to [scrape your data](#).

### Step 4: Choose your model(s)

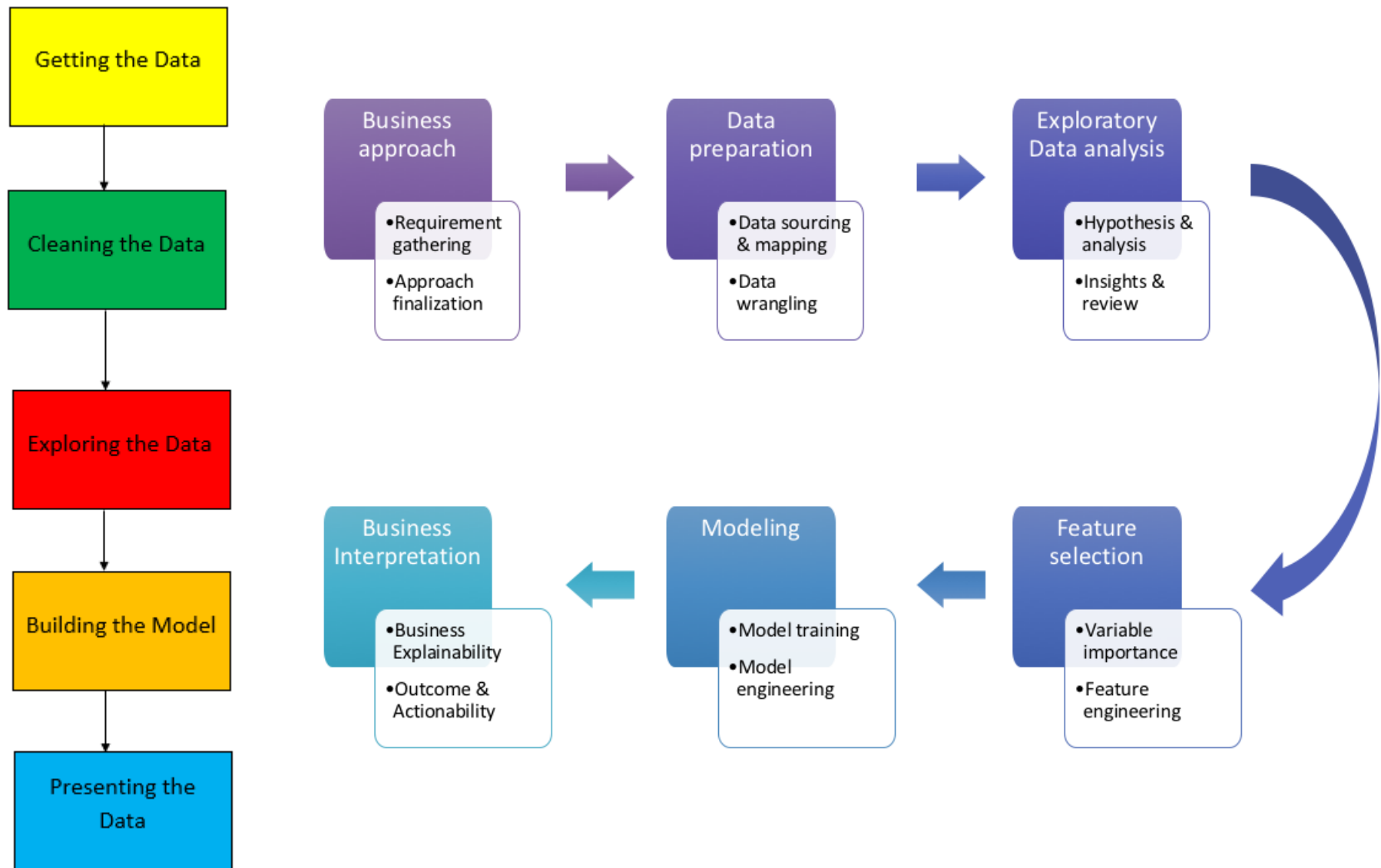
Choose your model(s) depending on your answers to these questions: are your outcomes discrete or continuous? Do you have [labeled or unlabeled datasets](#)? Are you concerned with outliers? How well do you want to interpret your results? The list of questions can vary depending on your project.

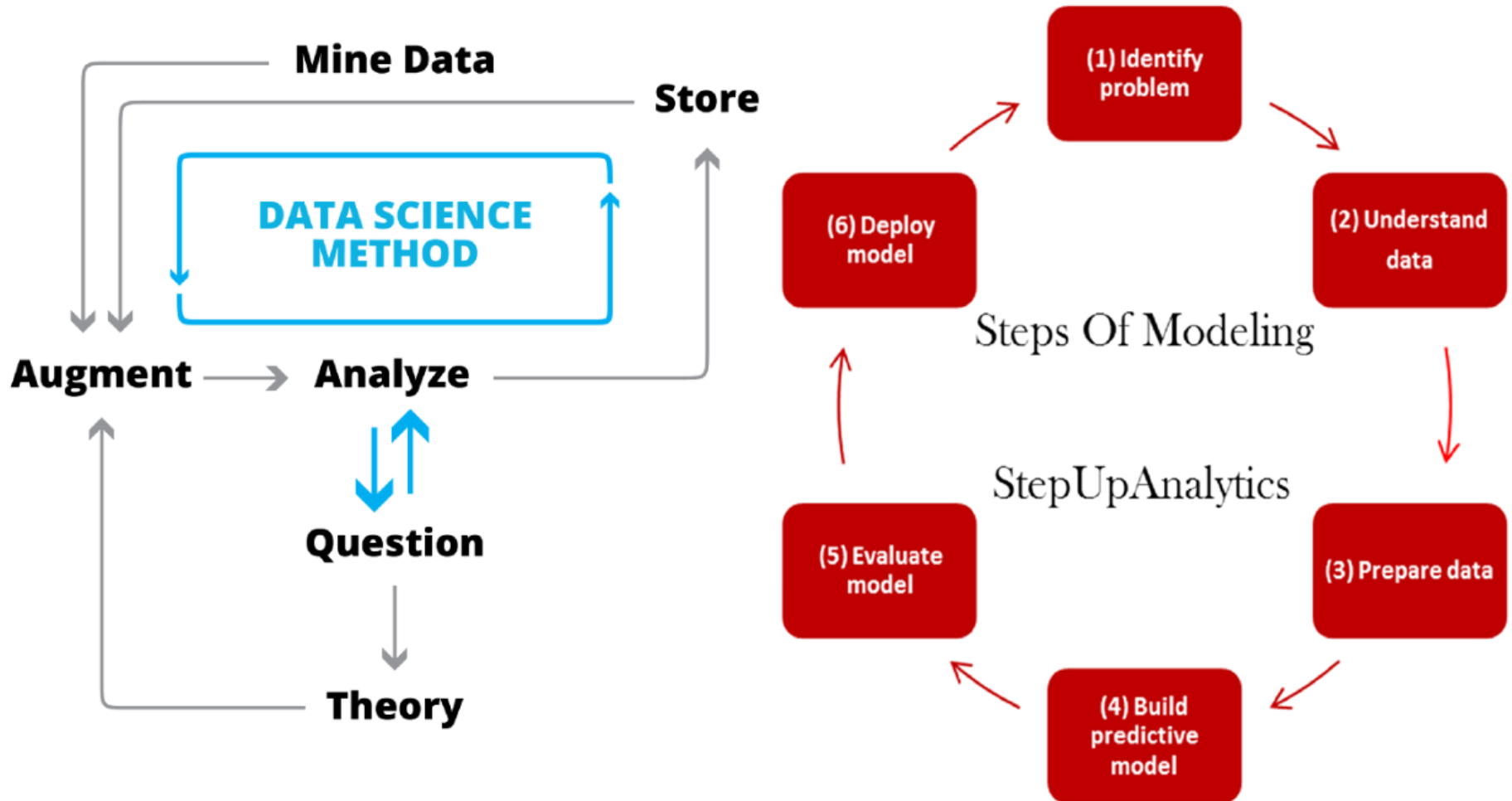
### Step 5: Identify model evaluation metrics

Identify corresponding model evaluation metrics to interpret your outcomes. Every model will have its own set of evaluation metrics.

### Step 6: Create a data preparation plan

What do you need to do to your data in order to run your model and achieve your outcomes? Data preparation includes data cleaning, feature selection, feature engineering, exploratory data analysis, and so on.





- For the project, form a team with 3 members
- Each team needs to pick one of the following data sets, propose several meaningful managerial questions that can be answered by the data set, use Python/R to analyze the data set and draw managerial insights.
- Grade criteria – propose/answer interesting managerial questions, build appropriate decision models, conduct clean data analysis via Python/R, and write an informative report!



- Each team needs to hand in a project proposal (8-10 pages) and give a short presentation (10-15 minutes) on **Apr 4**. The project proposal should include preliminary analyses, data descriptive statistics and data visualization of key variables.
- Each team needs to present their final results on **May 2** (30 minutes).
- The final report (30-40 pages) is due on **May 9 in class**.
- In the project proposal/report, each team needs to report the effort contribution of its team members (e.g., 30%, 30% and 40%). A student who makes less than 20% contribution will receive point deduction.

- Your project must be word-processed using a standard font (e.g., TimesNewRoman 12) and double-spaced.
- All charts, data sheets and tables (if any) should be done on the computer. Graphs and charts should have a title and properly labeled axes.
- All pages must be numbered.
- Order of pages:
  - COVER SHEET—The first page of your project must be a cover sheet that includes your name(s), date, the title of your project and **the percentage contribution of each team member's effort** .
  - EXECUTIVE SUMMARY - Briefly describe the problem, your solution and recommendation to managers.
  - TABLE OF CONTENTS



## Order of pages:

- INTRODUCTION, PROBLEM, PURPOSE—This section should include the statements of the project problem and goal.
- Data Sources: discuss your data details and properties.
- FINDINGS—Formulate the quantitative model and conduct quantitative analysis using Python/R.
- CONCLUSION—This is where you interpret your findings (what can you conclude or not conclude from your research) and recommend the solution to managers.
- REFERENCE—citations and websites referred if any.
- APPENDICES—attach anything technical or supplementary (e.g., spreadsheet models and explanations).

Send me the electronic version of your files, including Python/R codes, the project report and ppt presentation.

- Every team member should actively participate in the group project.
- In the project report, each team needs to report the effort contribution of its team members (e.g.,  $\frac{1}{3}$ ,  $\frac{1}{3}$  and  $\frac{1}{3}$  in a 3-member team).
- A student who makes less than 25% contribution will receive point deduction.
- If you feel that your teammate does not contribute much in the project, you should inform me immediately.

- A. LA Bike sharing – <https://bikeshare.metro.net/about/data/>
- B. Caterpillar pricing - <https://www.kaggle.com/c/caterpillar-tube-pricing/data>
- C. Bosch quality assurance - <https://www.kaggle.com/c/bosch-production-line-performance/overview>
- D. Airbnb (NYC, Berlin, Seattle and etc)  
<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/downloads/new-york-city-airbnb-open-data.zip/3>  
<https://www.kaggle.com/airbnb/seattle>  
<https://www.kaggle.com/brittabetendorf/berlin-airbnb-data>

- E. Capital Bike share – kaggle data  
<https://www.kaggle.com/marklvl/bike-sharing-dataset>
- F. Rossmann store sales – kaggle data  
<https://www.kaggle.com/c/rossmann-store-sales/data>
- G. Instacart shopping basket analysis  
<https://www.kaggle.com/c/instacart-market-basket-analysis>

- H. UK online retail data  
<http://archive.ics.uci.edu/ml/datasets/online+retail#>
- I. Black Friday sales <https://www.kaggle.com/mehdidag/black-friday/version/1>
- J. Retail chain store data  
<https://www.kaggle.com/manjeetsingh/retaildataset>
- K. Walmart store sales forecasting  
<https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>