MBAN 5110: PREDICTIVE MODELING

SESSION 1: INTRODUCTION TO PREDICTIVE MODELING

DR. ISIK BICER

TODAY’S AGENDA

- General course information
- Data Science
  - Challenges and opportunities
  - Sources of uncertainty
  - Classification of analytical approaches
- Prediction paradigm
- Mathematical Preliminaries

COURSE INFORMATION

- Instructor: Isik Bicer, PhD
- bicer@schulich.yorku.ca
- Please check Canvas regularly for course updates, announcements, and other information
- Course material:
  - Textbook: Biçer, Işık. *Supply Chain Analytics: An Uncertainty Modeling Approach*. Springer Nature, 2023.

COURSE TIMETABLE

Week	Date	Topic
1	Week 1	Introduction to Predictive Modeling
2	Week 2	Exploratory Data Analysis with Python
3	Week 3	Linear Regression
4	Week 4	Regularization Methods
5	Week 5	Dealing with Endogeneity Problems
6	**No Class**	**Thanksgiving**
7	Week 7	Classification Methods
8	Week 8	Time Series Analysis
9	Week 9	Unsupervised Learning
10	Week 10	Data-driven Prediction with Monte-Carlo Simulation
11	Week 11	Review
12	**No class**	**Final proposal + project coding (group)**

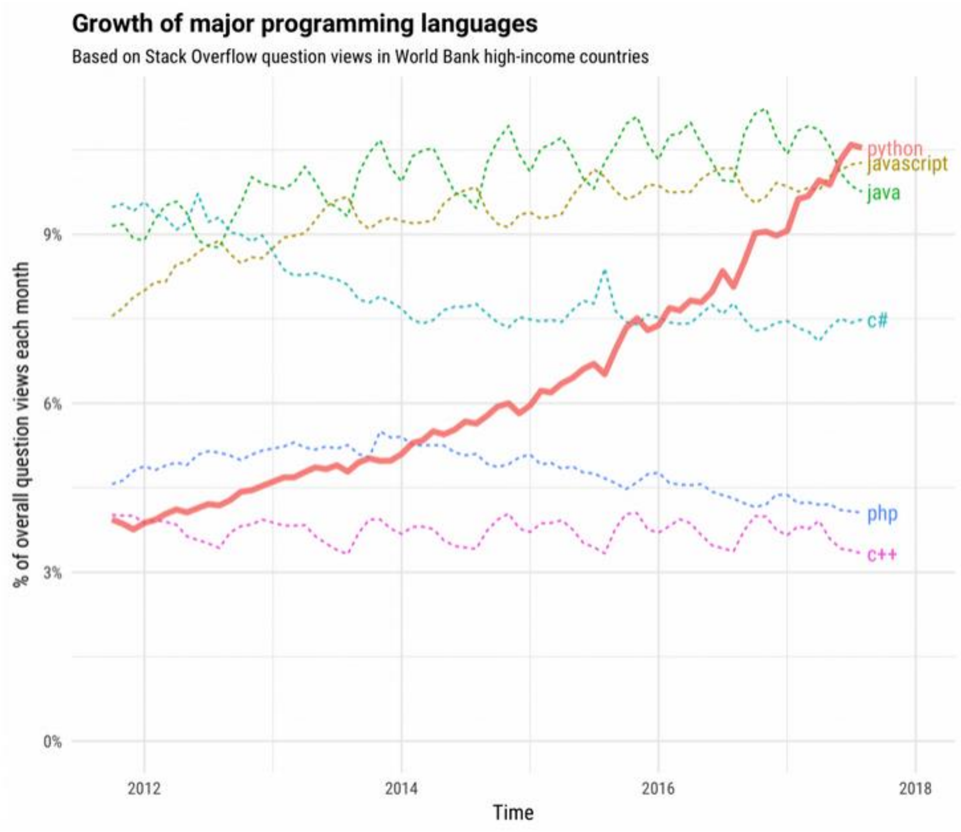
COURSE INFORMATION

- Grading

Assignment	Quantity	Weight	Total	
Assignment	1	25%	25%	Group
Midterm	1	30%	30%	Individual
Hackathon (Proposal)	1	10%	10%	Group
Hackathon (Coding)	1	35%	35%	Group

PYTHON

- The course emphasizes Python as a programming language for data science



Source: https://hackernoon.com/how-is-python-different-from-other-programming-languages-63311390f8dd

INTRODUCTION TO BUSINESS ANALYTICS AND DATA SCIENCE

DATA SCIENCE AND MACHINE LEARNING

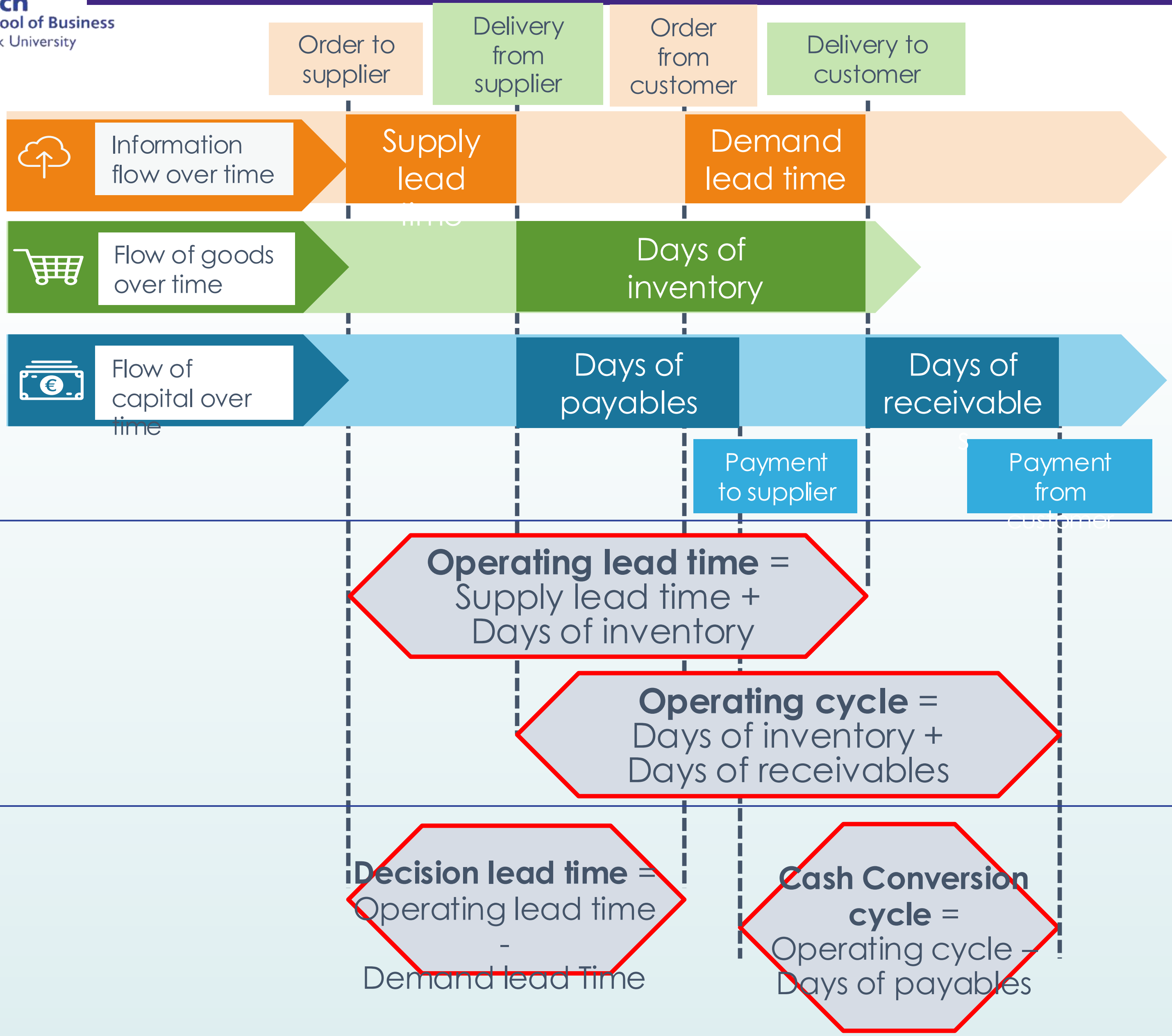
- Ultimate objective: To Learn from Data
- Why is it so important?
  - Complex business processes
  - Uncertainty about the future
  - Many sources of data that would alleviate the impact of complexity and uncertainty on our society
    - **“Every 2 days we create as much as information from the beginning of time until 2003”**
    - **“Google processes 1.2 trillion searches per year worldwide”**
    - **“Big Data could add 6 million jobs to US economy”**
    - **“Poor data analysis cost the US economy $3.1 trillion every year”**

Source: https://www.disruptordaily.com/7-facts-know-big-data/

INTERDISCIPLINARY STRUCTURE

- Discipline of turning data to useful insights with pragmatic validity and practical relevance
- Set of tools
  - Statistic
  - Linear Algebra
  - Optimization Theory
  - Computer Science
  - …

Sources of Uncertainty: Product Dashboard



DAYS	Max	Min	Avg.
Supply Lead Time	65	5	24
Demand Lead Time	42	14	22
Days of Inventory	95	40	43
Days of Payables	45	30	38
Days of Receivables	60	60	60

DAYS	Max	Min	Avg.
Operating Lead Time	97	35	45
Operating Cycle	152	100	103

DAYS	Max	Min	Avg.
Decision Lead Time	72	12	21
Cash Conversion Cycle	152	100	103

Sources of Uncertainty: Product Dashboard

How exposed is an organization to supply risk?

Risk indicator: Supply lead time

How exposed is an organization to process risk?

Risk indicator: Days of inventory

How exposed is an organization to demand uncertainty?

Risk indicator: Decision lead time = Operating lead time - Demand lead time

How exposed is an organization to cash flow uncertainty?

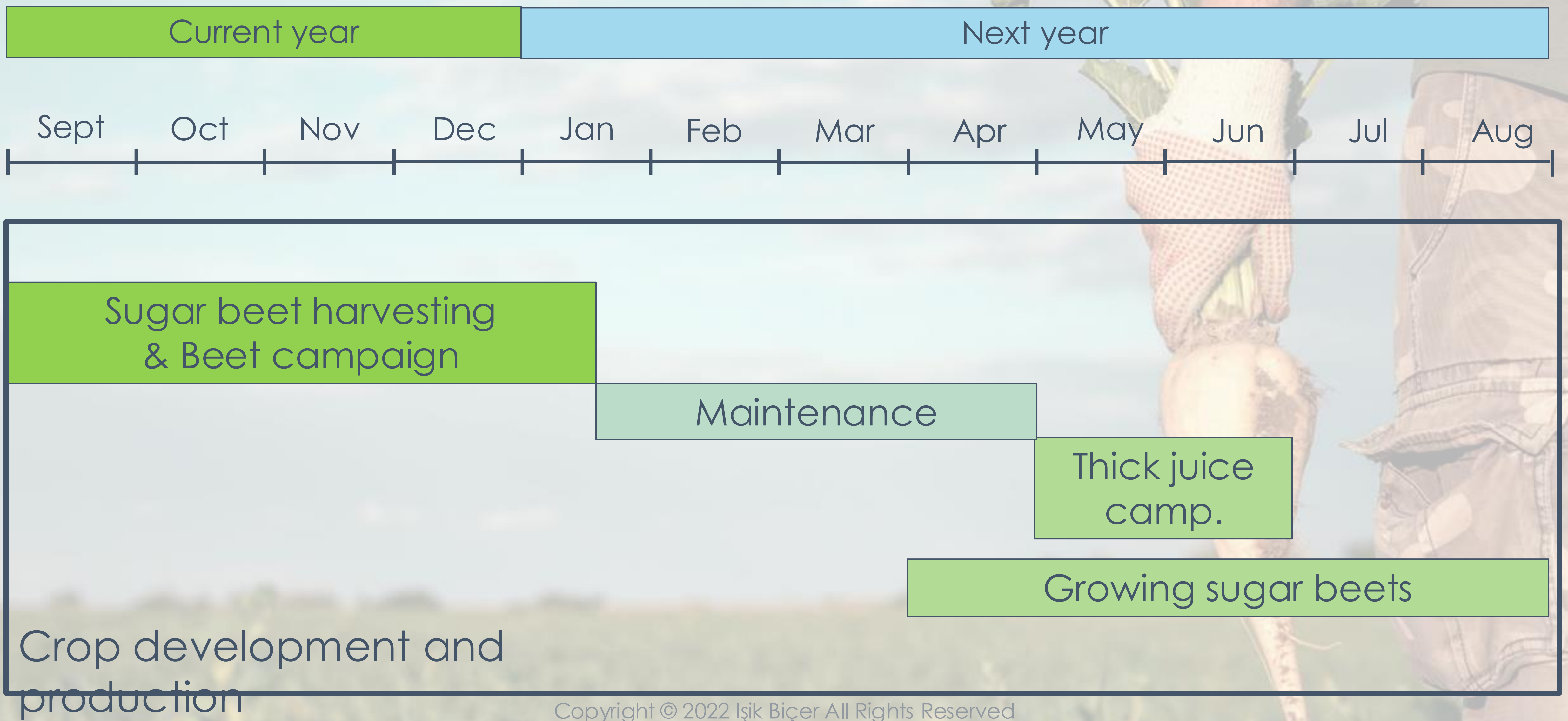
Risk indicator: Cash-conversion cycle = Operating cycle – Days of payables

DAYS	Max	Min	Avg.
Supply Lead Time	65	5	24
Demand Lead Time	42	14	22
Days of Inventory	95	40	43
Days of Payables	45	30	38
Days receivables	60	60	60

DAYS	Max	Min	Avg.
Operating Lead Time	97	35	45
Operating Cycle	152	100	103

DAYS	Max	Min	Avg.
Decision Lead Time	72	12	21
Cash Conversion Cycle	152	100	103

Uncertainty at the Supply Side



Uncertainty at the Supply Side

Sugar supply is uncertain:
Combination of three uncertain elements

#1 Uncertainty:
Total harvest area

Farmers are incentivized to grow sugar beets, but not forced

#2 Uncertainty:
Sugar beet amount per hectare

Amount of sugar beets per hectare is affected by climate factors during the growth period

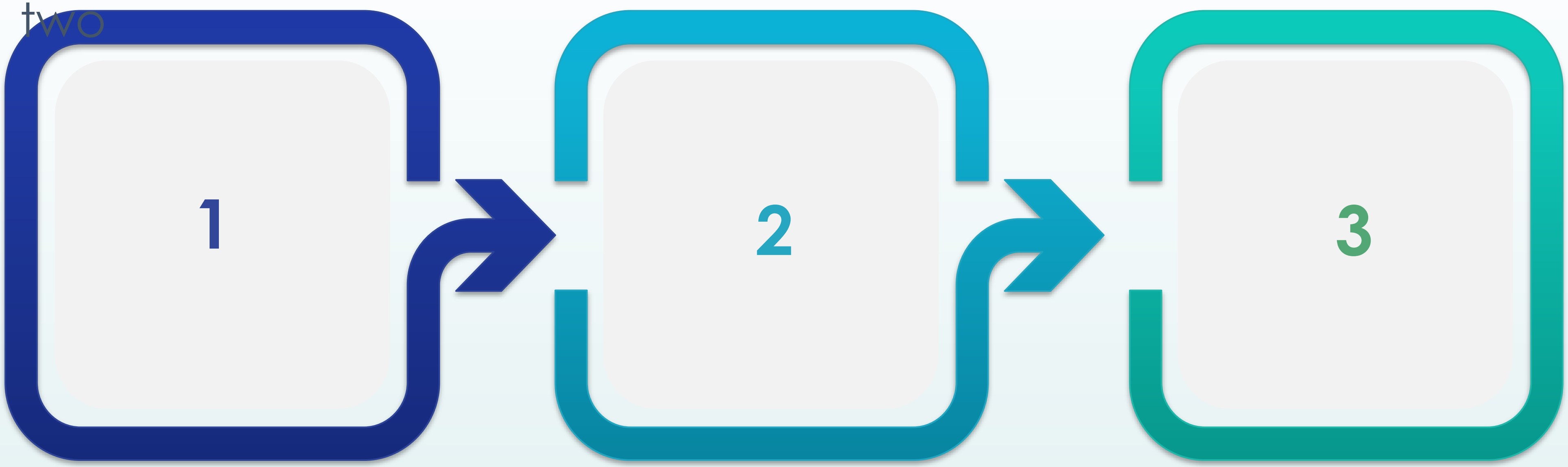
#3 "Uncertainty:
Percentage sugar content
per kg of sugar beets

Sugar content per one kg of sugar beets is affected by climate factors during the harvesting period

Uncertainty at the Operation Side

Serial production with three machines

- Failure rates are uncertain for each machine.
  - First machine is less reliable than the other two



Total amount of production is uncertain:
Combination of three uncertain elements that come from each machine

Uncertainty at the Demand Side

Kordsa receives bulky orders from its customers

1. The number of distinct orders in a week is uncertain
2. For each order, the quantity demanded is uncertain
3. For each order, demand lead time is uncertain

ANALYTICAL APPROACHES TO BUSINESS PROBLEMS

Descriptive Analytics

- *What happened?*
- *Why did it happen?*

Predictive Analytics

- *What will happen?*

Prescriptive Analytics

- *How can it be improved?*

Technical difficulty →

OBJECTIVES OF DESCRIPTIVE ANALYTICS

Schulich School of Business York University

- Describe the business context
- Develop some performance metrics and monitor them periodically
- Identify key parameters that have a strong influence on performance metrics
- Identify business opportunities that can be done without difficulty but have the strong positive impact on the business

OBJECTIVES OF PREDICTIVE ANALYTICS

Schulich
School of Business
York University

- Model uncertainty
  - There is always uncertainty in business contexts. In the retail industry, demand is uncertain. In the finance industry, stock prices are uncertain
- Minimize reducible variance of the model
- Minimize the model bias
  - The last two objectives may not be attainable at the same time. ***There is a variance-bias trade-off***

OBJECTIVES OF PRESCRIPTIVE ANALYTICS

Schulich
School of Business
York University

- Support decision-making process in organizations
- Decision makers in corporations make sequential and interlocking decisions every day under uncertainty given some constraints
- Optimization models are effective in casting real world problems into mathematical frameworks and solve them accordingly

DATA MANAGEMENT

DATA COLLECTION

- Data can be either primary or secondary
- Primary data:
  - Researchers collect the primary data themselves for specific purposes
  - Surveys, interviews, direct observation
  - E.g., Darwin’s data collection on his South America visit on HMS Beagle (1831-1836)
- Secondary data:
  - Data collected routinely as a part of operations of a company, institution, etc.
  - E.g., Census data

DATA COLLECTION

- Primary data:
  - In general, it is more expensive to obtain
  - Online surveys may be cheaper but not effective and reliable most of the time
  - Not needed for daily decision making
- Secondary data
  - Less expensive once the system to collect the data is constructed
  - Needed for daily decision making more often than primary data
  - Usually requires a preprocessing to bring the data into the requested form

DATA MANAGEMENT

- Includes all the activities starting from Design of Data Architecture to Data Governance
  - Step 1: Design of Data Architecture
  - Step 2: Data Normalization
  - Step 3: Data Processing
  - Step 4: Data Governance

DATA ARCHITECTURE

- Databases are complex designs that include metadata, transactional data, their relationships.
- Data architects deal with the questions what types of metadata are needed, how transactional data is handled
- E.g., Sales records in autodealer
  - Transactional data: Each sale record
  - Meta data: Cars, sales representatives

AUTO DEALER EXAMPLE

<u>Sales Records</u>
Unique identifier
Date of sale
Car Sold
Price
Sales Representative

Metadata

<u>Car Inventory</u>
Car ID
Model
Year
Color
Type (SUV, Hatchback …)

<u>Employee</u>
Employee ID
Name and Surname
Job Title

DATA NORMALIZATION

- One of the main principles of data management is that nulls should be avoided.
- If a cell can be null, the column should be stored in a separate data table because the data tables should be complete.
- We call this process normalization.
- Suppose that the “Car Sold” and “Sales Representative” information are missing for some transactions in the Sales Records datatable.

AUTO DEALER EXAMPLE

Not Normalized

Sales Records
Unique identifier
Date of sale
Car Sold (NULL)
Price
Sales Representative (NULL)

Normalized

Sales Records
Unique identifier
Date of sale
Price

SalesRepresentative Table
Unique identifier
Sales Representative

SalesCar Table
Unique identifier
Car Sold

DATA PROCESSING

- Retrieving data through SQL queries and transforming it to different forms
  - E.g., we may be interested in which models in the autodealer are on high demand. So, we may use SQL with group by and sum functions to bring the sales records data into the desired form.
- Integrating data from various sources: Inaccurate and incomplete data is detected first. Then, such records are modified or replaced. We call this process "*Data Cleansing*"
  - Next week: Data imputation techniques in Python

DATA GOVERNANCE

- Who can reach the data?
- Is data sensitive or not?
- What protocols should be implemented in case of data misuse?
- How to protect data from cyberattacks?

PREDICTION PARADIGM: VARIANCE-BIAS TRADE-OFF

THE PREDICTION PARADIGM

$$y_i = f(x_i) + \epsilon_i$$

- Output variable: y_i
- Input variable: x_i
- Random error term: ϵ_i such that $E(\epsilon) = 0$ and $Var(\epsilon) = E(\epsilon^2) - E(\epsilon)^2 = \sigma^2$
- $f(\ldots)$: True model
- Suppose we don't know the true model but use an estimate
- $g(\ldots)$: Estimate model such that our prediction for the output becomes

THE PREDICTION PARADIGM

$$y_i = f(x_i) + \epsilon_i$$

- What happens to the mean squared error if we use the estimate model:

$$E(y_i - g(x_i))^2 = E(f(x_i) + \epsilon_i - g(x_i))^2$$
$$= E(f(x_i) - g(x_i))^2 + 2E(\epsilon_i)E(f(x_i) - g(x_i)) + \sigma^2$$

- The second term is zero because $E(\epsilon_i) = 0$

$$E(y_i - g(x_i))^2 = \underbrace{\boldsymbol{E(f(x_i) - g(x_i))^2}}_{\text{Reducible modeling variance}} + \underbrace{\boldsymbol{\sigma^2}}_{\text{Irreducible variance}}$$

THE BIAS-VARIANCE TRADE-OFF

- Let's focus on the reducible modeling variance

$$E(f(x_i) - g(x_i))^2 = E(f(x_i) - g(x_i) + E(g(x_i)) - E(g(x_i)))^2$$
$$= E(f(x_i) - E(g(x_i)))^2 + E(g(x_i) - E(g(x_i)))^2$$

- Be aware that

$$2E\big(f(x_i) - E(g(x_i))\big)E\big(g(x_i) - E(g(x_i))\big) = 0$$

- The bias-variance trade-off function is:

$$E(f(x_i) - g(x_i))^2 = \underbrace{E(f(x_i) - E(g(x_i)))^2}_{\text{Square bias of } g(x)} + \underbrace{E(g(x_i) - E(g(x_i)))^2}_{\text{Variance of g}(x)}$$

THE BIAS-VARIANCE TRADE-OFF

- Bias is the difference between the true model prediction and the average prediction of the estimate model
- Variance is the average variability of the prediction of the estimate model
  - The variance exists because coefficients of the estimate model $g(x_i)$ would change when training set is updated
  - The variance also increases as the model becomes more complex

THE BIAS-VARIANCE TRADE-OFF

- The variance also increases as the model becomes more complex

$$g(x_i) = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \cdots$$

- True model is much more complex than the estimate model
- Each coefficient is an estimate that also has a variance
- When we resample training set, the estimate would have a different value within some confidence intervals

$$Var(g(x_i)) \\ = Var(a_0) + Var(a_1)x_{i1}^2 + Var(a_2)x_{i2}^2 + Var(a_3)x_{i3}^2 + \cdots$$

MATHEMATICAL PRELIMINARIES: LINEAR ALGEBRA

ROLE OF LINEAR ALGEBRA IN DATA SCIENCE

- Data tables are used in the matrix form
- Classical sum-product formulations are not computationally efficient
- Analytical applications should be maintained while preserving the matrix form of data
  - Data management practices are based on this principle

MATRIX OPERATIONS

- Consider this expression:

$$Ax = b$$

where

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}, \qquad A = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & & a_{2k} \\ \vdots & \ddots & \vdots \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix}$$

- x vector of unknowns

INVERSE OF A MATRIX

- A^{-1} is the inverse of matrix A

$$A^{-1}A = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Identity matrix I is a diagonal matrix that has only ones in diagonal cells

$$Ax = b$$
$$A^{-1}Ax = x = A^{-1}b$$

- So, we can find a solution of x if we can calculate the inverse of A

MATRIX MULTIPLICATION

- Inner product: $\mathrm{Ax} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 \\ 2x_1 + 5x_2 + x_3 \\ x_1 + 2x_2 + 4x_3 \end{bmatrix}$
- The inner product is also known as “dot product”. The first row is obtained by taking the dot product of:

$$[1 \quad 3 \quad 0].[x_1 \quad x_2 \quad x_3] = 1x_1 + 3x_2 + 0x_3$$

- Not so interesting because we destroy the columns
- How can we preserve the column structure while doing the matrix multiplication

MATRIX MULTIPLICATION

- Preserving the column structure by rewriting the multiplication as follows:

$$\begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 \\ 2x_1 + 5x_2 + x_3 \\ x_1 + 2x_2 + 4x_3 \end{bmatrix}$$

- Ax is a linear combinations of the columns

COLUMN SPACE

$A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ and write each column as follows

$$A_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, A_2 = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}, A_3 = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

- Every combination of $A_1x_1 + A_2x_2 + A_3x_3$ is in the column space
- $A_1x_1 + A_2x_2 + A_3x_3 = b$ has a unique solution for any b so the column space R^3
- Thus, A is an invertible matrix such that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = A^{-1}b$

COLUMN SPACE

Let's change the matrix as $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 3 \\ 1 & 2 & 1 \end{bmatrix}$ and write each column as follows

$$A_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, A_2 = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}, A_3 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

- Now, $A_3 = A_2 - A_1$ so A_3 is dependent
- "$A_1x_1 + A_2x_2 + A_3x_3$" can be written as "$A_1\widehat{x_1} + A_2\widehat{x_2}$"
- $Ax = b$ for $b = \begin{bmatrix} 0 \\ 0.25 \\ 1 \end{bmatrix}$ doesn't have any solution.

BASIS OF A MATRIX

- Basis: Subset that involves only independent columns
  - Column space of R^3

$$A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix} \rightarrow C = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

Rank of the matrix is 3

  - Column space of a plane

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 3 \\ 1 & 2 & 1 \end{bmatrix} \rightarrow C = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 1 & 2 \end{bmatrix}$$

Rank of the matrix is 2

  - Column space of a line

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix} \rightarrow C = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Rank of the matrix is 1

FORMING THE BASIS: COLUMN SPACE

- Put the first non-zero column to the matrix
- Put the second non-zero column if it is not the multiplication of the first one
- Put the third non-zero column if it is not the combination of the previous ones

ROW SPACE

- The operations can be repeated for the rows to construct the row space.
- The number of independent columns is equal to the number of independent rows

DECOMPOSING THE MATRIX INTO INDEPENDENT COLUMNS AND ROWS

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix} \rightarrow \mathrm{C} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

Rank of the matrix is 1

- $\mathrm{R} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$
- $A = CR = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix}$

DECOMPOSING THE MATRIX INTO INDEPENDENT COLUMNS AND ROWS

- $A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 3 \\ 1 & 2 & 1 \end{bmatrix} \rightarrow \mathrm{C} = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 1 & 2 \end{bmatrix}$ ***Rank of the matrix is 2***
- $\mathrm{R} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$
- $A = CR = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 5 & 3 \\ 1 & 2 & 1 \end{bmatrix}$

DECOMPOSING THE MATRIX INTO INDEPENDENT COLUMNS AND ROWS

Schulich School of Business York University

- $A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix} \rightarrow \mathrm{C} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix}$

Rank of the matrix is 3

- $\mathrm{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- $A = CR = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 5 & 1 \\ 1 & 2 & 4 \end{bmatrix}$

DECOMPOSITION IN THE OUTER PRODUCT FORM

- $\begin{bmatrix}1 & 3\\ 2 & 5\\ 1 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & 1\end{bmatrix} = \begin{bmatrix}1\\ 2\\ 1\end{bmatrix}\begin{bmatrix}1 & 0 & -1\end{bmatrix} + \begin{bmatrix}3\\ 5\\ 2\end{bmatrix}\begin{bmatrix}0 & 1 & 1\end{bmatrix} =$

First column and first row

Second column and second row

$$\begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix} + \begin{bmatrix}0 & 3 & 3\\ 0 & 5 & 5\\ 0 & 2 & 2\end{bmatrix} = \begin{bmatrix}1 & 3 & 2\\ 2 & 5 & 3\\ 1 & 2 & 1\end{bmatrix}$$

WHEN THE MATRIX IS NOT INVERTIBLE

- Suppose A is an $n \times n$ matrix

$$\underbrace{A = \begin{bmatrix} & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \end{bmatrix}}_{\text{n columns}} \left.\vphantom{\begin{bmatrix} \\ \\ \\ \end{bmatrix}}\right\} \text{n rows}$$

- If some of the columns are not independent, the matrix is not invertible
- Rank_of_$A < n$: such a matrix is singular
- To be invertible, a square matrix should be a full rank matrix:
  - Rank_of_$A = n$