

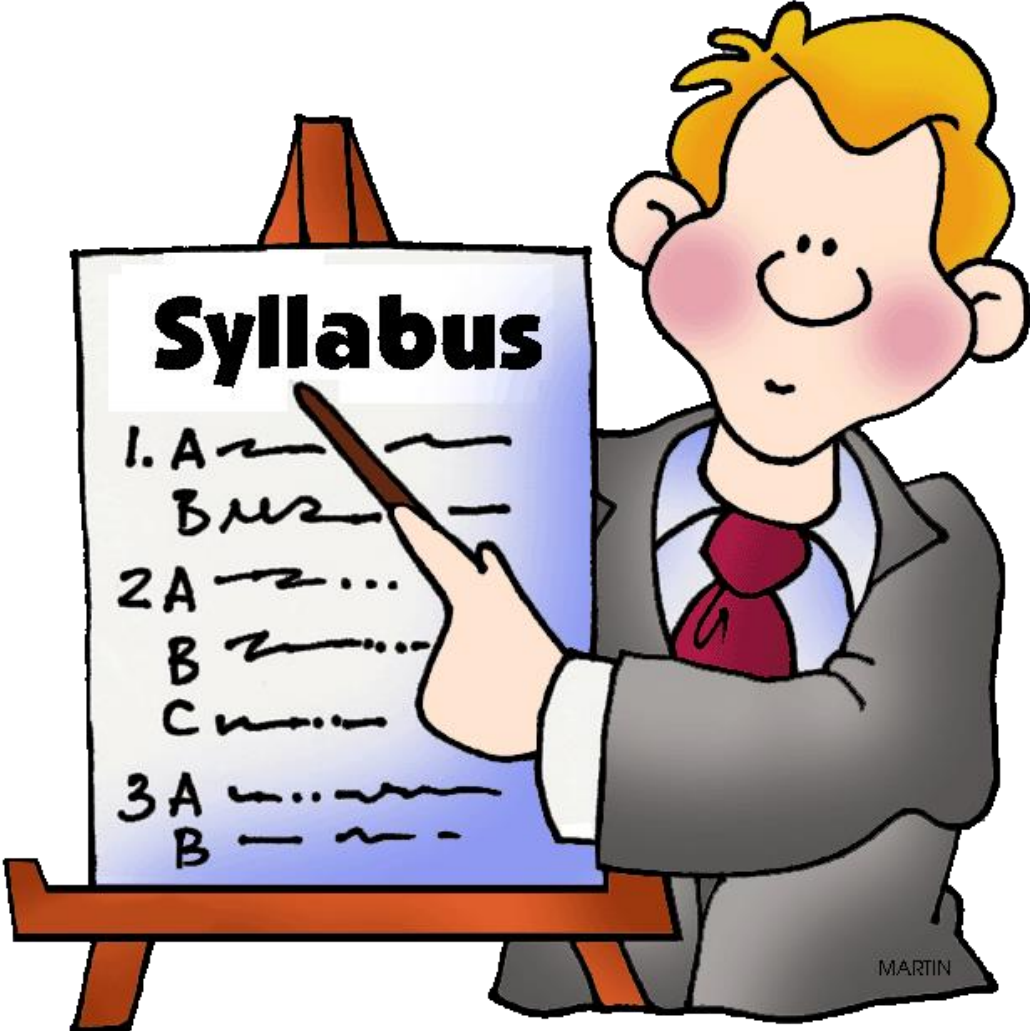
**Hacettepe University**  
**Department of Industrial Engineering**  
**Undergraduate Program**

**EMU 430 – Data Analytics**

2023-2024 Fall  
Friday 13:40 – 16:30

**Instructor:** Erdi Dasdemir

[edasdemir@hacettepe.edu.tr](mailto:edasdemir@hacettepe.edu.tr)  
[www.erdidasdemir.com](http://www.erdidasdemir.com)



# Introduction to Data Analytics

Introduction Three Types of Analytics Prescriptive Analytics OR Data Language 4/48

# Three Types of Analytics

Introduction Three Types of Analytics Prescriptive Analytics OR Data Language 12/48

# Prescriptive Analytics: Operations Research

Introduction Three Types of Analytics Prescriptive Analytics OR Data Language 20/48

# Programming and R

Introduction Three Types of Analytics Prescriptive Analytics OR Data Language 29/48

# Introduction to Data Analytics

- Focus on descriptive analytics and to develop practical skills for applying data analytics to explore raw data.
- Learn fundamental programming skills for data analytics and hands-on skills in working with R language.
- This course is for everybody:
  - This course expects you to be comfortable with fundamental mathematics, algebra, and logical operations.
  - coding experience is not necessary

○ We will learn

- to collect and transform raw data in various formats,
- utilize advanced visualizations to draw actionable insights
- use productivity tools to organize data analytics projects and create reproducible reports.

# What is Analytics?

○ Analytics refers to using math, statistics, operations research, computer programming and computational resources systematically to transform data into actionable insights for making better decisions.



○ Convert raw data into a rich story that you can act on.

# What is Data Analytics?

- "Analytics" and "data analytics" are closely related terms, and their meanings can overlap.
- Main focus is on handling raw data
- Gather, process, interpret, and present data





- Humans are naturally talented at using analytics to make informed decisions.
- For example, don't we all adjust the time it takes us to cook an egg in the morning based on our previous experience or change the route we use to go to work or school when there is traffic on our regular road?
- If it's that easy to make an informed decision, **why do we need data analytics?**



# Why do we need data analytics?

- **Real life and business problems are often bigger.**
  - For example, we can't immediately answer questions like "*What is the right time to have a child for me?*" or "*Which car should I buy with my limited budget?*".
  - We have to process a lot of data, consider many criteria, evaluate hundreds of different solutions, and select the best one.

# DATA SCIENTIST



What my friends think I do



What my mum thinks I do



What society thinks I do



What my boss thinks I do



What I think I do



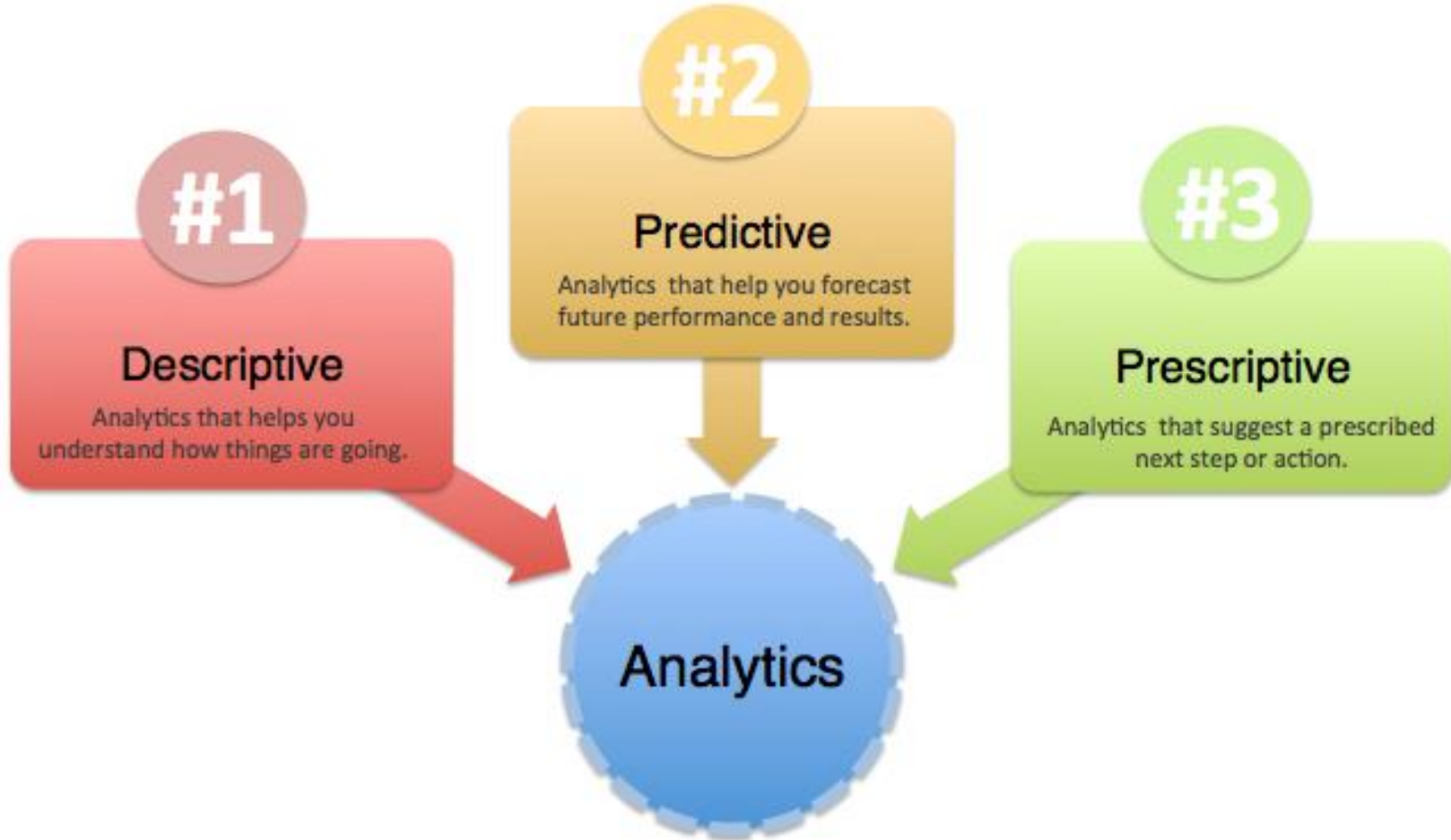
What I actually do

If you're in a "data scientist" role in your business life, you're likely to be asked something like

*"Can you estimate our sales potential for the next year by studying our last 10 years of demand data? It would be great if you also determine the optimum production capacity of the new factory we will open."*

# Three Types of Analytics

# 3 Types of Analytics



<https://gblogs.cisco.com/uki/analytics-what-next>

○ ***Descriptive Analytics*** expertise in

➤ summarizing,

➤ visualizing,

➤ and thus uncovering the performance of a system

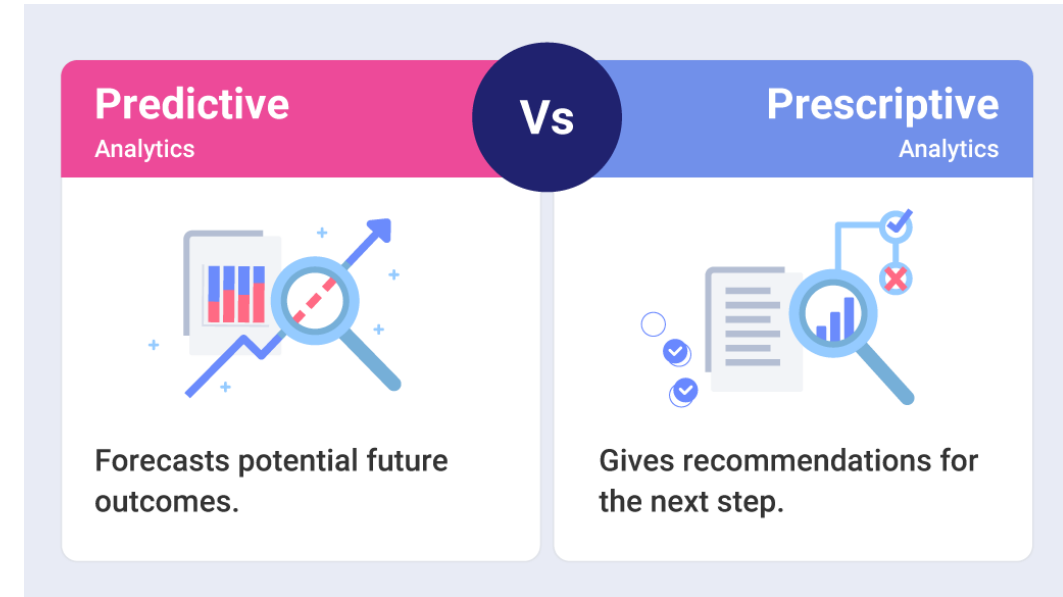
from data that cannot be interpreted through manual calculations and observations.

- ***Predictive Analytics*** expertise in developing models that use existing data to predict future outcomes.
- The dominating field of study for this area is?
  - **Machine learning studies are the pioneers of this field**

- **Prescriptive Analytics** expertise in developing models that use past data to generate prescriptive results to decision makers.
- Generating prescriptive results means?
  - **providing certain recommendations for decision problems**
- The dominating field of study for this area is?
  - **Operations Research**

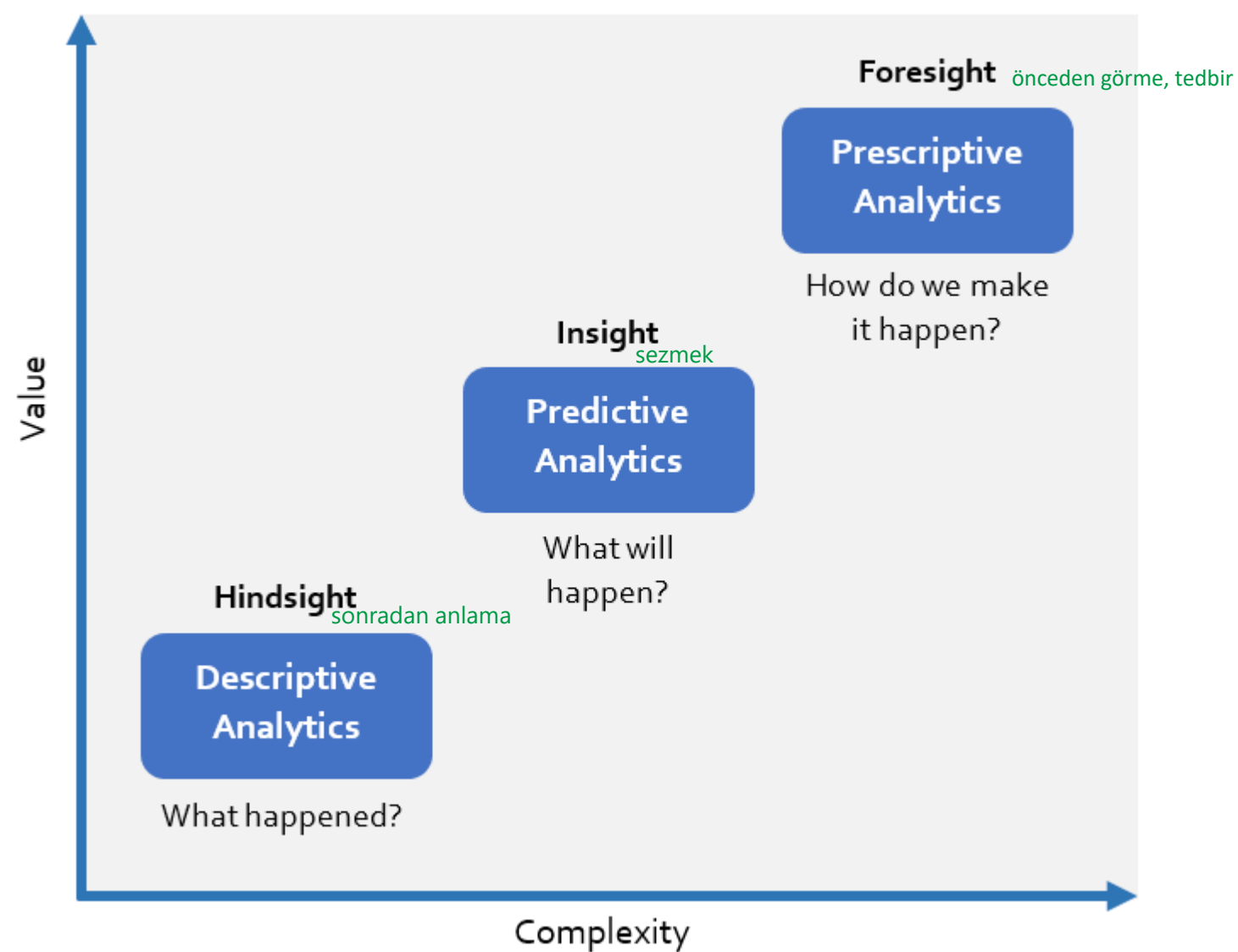


- **Prescriptive Analytics** differ from the previous two analytics in that
  - they can provide the decision maker with a ready-made decision
  - prove that it is the “best” one.



<https://www.linkedin.com/pulse/descriptive-predictive-prescriptive-analytics-2023-marc-mandel-ccxp/>

- It is gaining increasing popularity and is also well suited to the background of **Industrial Engineering** students.



<https://zyabkina.com/descriptive-predictive-prescriptive-analytics-in-marketing/>

- Operations Research Analyst - The fastest growing job you have never heard of (2021)

- Increasing popularity

- ease of access to data
- advancements in computer programming
- computing technologies



# Prescriptive Analytics: Operations Research





Prof. Dr. Russell Lincoln Ackoff  
1919-2009

Father of  
Operations Research



Prof. Dr. Halim Doğrusöz  
1922 – 2019

Father of  
Operations Research in Turkey

RUSSELL LINCOLN ACKOFF'U SEVGİ VE SAYGIYLA ANIYORUZ

Halim Doğrusöz

2010

**RUSSELL LINCOLN ACKOFF'U SEVGİ VE SAYGIYLA  
ANIYORUZ**

*Halim DOĞRUSÖZ*

*Endüstri Mühendisliği Bölümü, Orta Doğu Teknik Üniversitesi, 06531 Ankara  
halim@ie.metu.edu.tr*

*Geliş Tarihi: 15 Ocak 2010*



Prof. Dr. Russell Lincoln Ackoff  
1919-2009

Father of  
Operations Research

*“Russ’la, önce hoca - öğrenci ilişkisi olarak başlayan, sonra sıkı bir dostluğa dönüşen bir beraberliğimiz oluşmuştur. İlişkimiz bana Case OR GROUP’ta bir yer vermesiyle başlar. Bu bana GROUP’ta ilk ve tek Türk araştırma asistanı ayrıcalığını bahşetmiştir. Aramızda sadece üç yıllık bir yaş farkı var. Ben 1958’de GROUP’a katıldığım zaman 36 yaşında bir yaşlı asistan o ise 39’ında bir genç yönetici idi. Sonra iş beraberliğimiz bizi yakınlaştırdı. Bu yakınlığımız onun Türkiye’ye de yakınlığını geliştirdi.*

....

*Yakınlığımız nedeniyle, Russ’ın Türkiye’ye ilgisi vardı; ama ben vatana dönüp çalışmalara başladıktan sonra, tabiatıyla, bu ilgi çok daha arttı. Bu şekilde kendisinin fikirlerinden yararlanma olanağı doğdu. Birçok kez Türkiye’ye çağırdık ve 1960’ların ikinci yarısında hiç nazlanmadan beş kez geldi ve birçok konuda önemli öneri ve yönlendirmeleri oldu.”*





Prof. Dr. Russell Lincoln Ackoff  
1919-2009

Father of  
Operations Research

“... ”

*Elbette “focus” YA’nın Türkiye’de gelişiminde ve bunun için de Ünite’nin gelişiminde idi. Ama Ünite TÜBİTAK’ın Ünitesi ve TÜBİTAK da yeni kurulmuş gelişmekte olan bir kurumdu. Dolayısıyla TÜBİTAK da çalışmaların kapsamındaydı.*

*Her gelişinde Ünite’nin konuları konuşulurken TÜBİTAK’ın programları, projeleri, politikaları, planları da gündeme geliyordu. Bilim Kurulu Üyeleri, özellikle Bilim Kurulu Başkanı Cahit Arf, Genel Sekreter Nimet Özdaş ve diğer yöneticilerle toplantılar yapıyor görüş teatisinde bulunuyorduk. Tabiatıyla Russ da görüşlerini belirtiyor önerilerde bulunuyordu. Doğal olarak sorunların çözümünde Ünite’nin rolü ve katkısı da konuşuluyordu.*

...”



In Memoriam: Professor Halim Doğrusöz (1922-2019)

Murat Köksalan and Haldun Süral



Prof. Dr. Murat Köksalan



Prof. Dr. Halim Doğrusöz  
1922 – 2019



Prof. Dr. Haldun Süral

“ ...

*Profesör Doğrusöz, Türkiye'deki ilk YA araştırma grubunu TUBİTAK'a bağlı olarak 1965'te kurdu. Orta Doğu Teknik Üniversitesi'nde ilk YA yüksek lisans programını (1965) ve ilk YA ve İstatistik lisans programını (1975) kurdu. Bir üniversitenin temel sorumluluklarından birinin, toplumun sorunlarını araştırma yoluyla ele alarak topluma hizmet etmek olduğuna inanıyordu.*

...

*Profesör Doğrusöz'ün YA yaklaşımını özgünlüğüyle uygulama tutkusu Türk akademik çevrelerinde büyük ses getirdi. Uygulamalı araştırma yürütme konusundaki kararlılığı bulaşıcı oldu. Bunun muhtemelen YA'nın Türkiye'deki şaşırtıcı derecede popüler olmasında önemli bir etkisi oldu.”*



## Yöneylem Araştırması/Endüstri Mühendisliği “YA/EM” Kongreleri

- İlki “Ulusal YA Kongresi” adıyla 1975 yılında Boğaziçi Üniversitesi
- Toplam 41 Ulusal YA/EM kongresi ile üniversite, kamu, sanayi ve uygulayıcıları bir araya gelmiştir.
- 42. Yöneylem Araştırması ve Endüstri Mühendisliği Ulusal Kongresi
  - <https://www.yaem2023.org/>
  - 1-3 Kasım 2023, Gaziantep (Gaziantep Üniversitesi ve Hasan Kalyoncu Üniversitesi ortaklığında)
  - "Sürdürülebilirlik için YA/EM"



2017, İstanbul



2018, Eskişehir

### Operations Research

- "Research" conjures up the scientific method, i.e. "research" as a precursor to "practice."

### Mathematical Programming

- makes many people think of computer science.

### Optimization

- very broad term  
- search engine optimization or code optimization. □ useful tools in their own right, they're relegated to very narrow segments of business practices,

### The Science of Better

- 2000s, the professional society in this area (INFORMS)

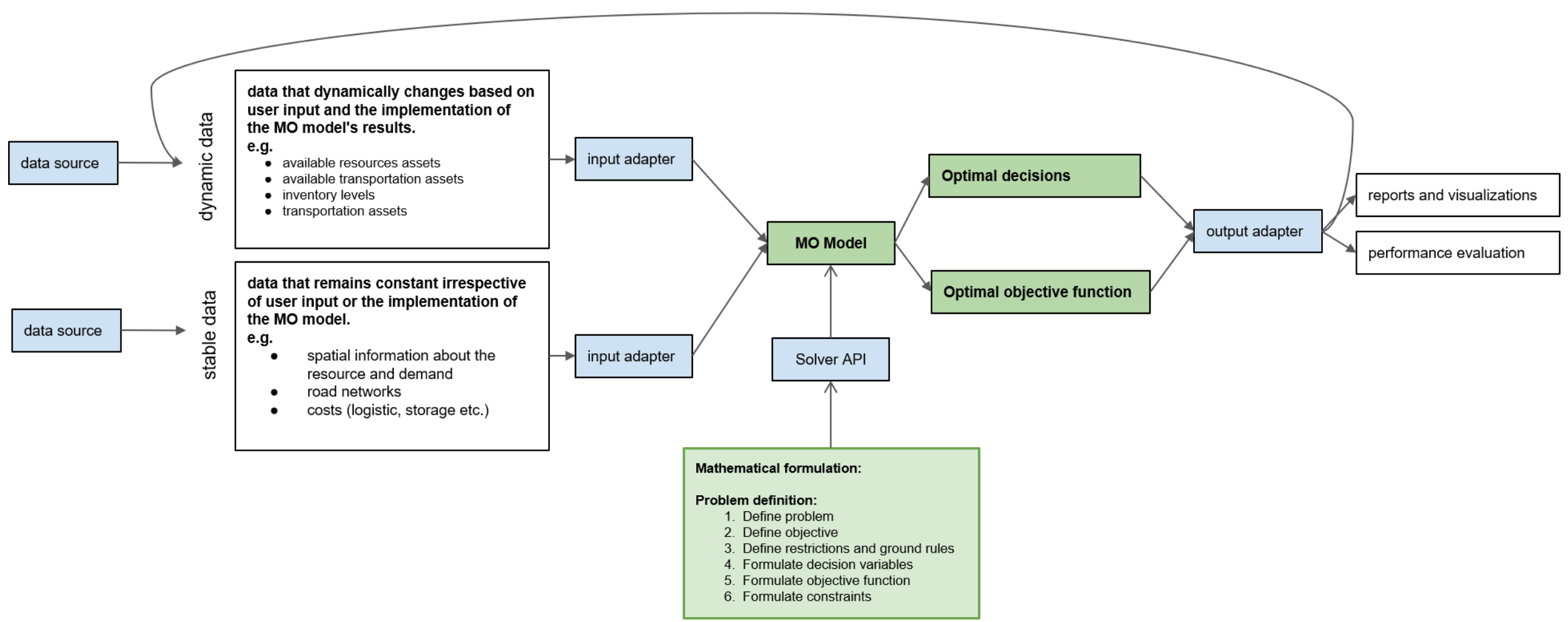
### Mathematical Optimization (Matematiksel Optimizasyon)

- use mathematics to solve business problems.  
- the process of finding values that optimize a mathematical function  
- lead by GUROBI

**History of the Name(Edward Rothberg ,2022, Mathematical Optimization: What's In A Name?, FORBES)**



# Mathematical Optimization in Software Projects



# Programming and R

- **Why do we need computer programming for data analytics?**
  - Technology has become so pervasive that it is not just the product of engineers, but of everyone.
  - Law offices, healthcare systems in hospitals, public transportation tracking systems, and even the pedometers in our smartwatches, all rely on technology
  - Everyone today is a natural data generator; even my baby generates data, as we measured her temperature, recorded her feeding durations, and kept track of the number of diapers we change.

- **Why do we need computer programming for data analytics?**
  - To survive in such a world, the best strategy is to position ourselves on the side of the creators of technology.
  - When robots take over the world, it may be beneficial to be among those who created them, with the assumption that they will show respect to their creators!

- **Why do we need computer programming for data analytics?**
  - Cannot we use just software like Excel and Minitab?
  - Unfortunately, the answer is no. Why?
  - These programs are great, but the possibilities are often limited by their creators. Programming, however, provides us with limitless potential



- Learning a programming language initially can be awkward and challenging.
- Computers may seem smart, but they don't always understand what you're saying; you need to be precise.
- However, the most difficult part is the beginning. Once you're familiar with it, your progress will be very fast

- Additionally, popular programming languages of data science, R and Python, have large communities with answers to almost any question you may have. I can't remember a time when I couldn't find the answer online.
- If you have a question that doesn't have an answer on the web, you may be about to uncover something new and exciting!

○ R programming language

- We will be using the R programming language in this course.
- The other popular and wonderful data science languages are Python and, more recently, Julia, which is gaining popularity day by day.
- Although we will use R in our code samples, all data analytics tasks we will cover can be done with Python or Julia with the same effort.
- Once you feel comfortable with these languages, you will notice that you don't need to choose one of them; in fact, using them all together will give you more power.

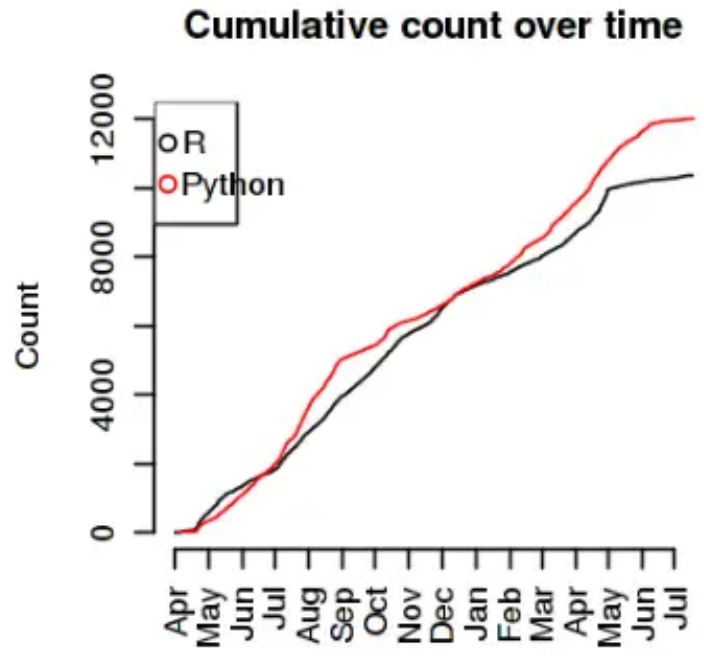


Figure 7. Software used in data science competitions on Kaggle.com in 2015 and 2016.

- **R programming language**

- Open-source programming language
- Some refer to it as statistical language due to its origin
- Grew out of Bell Labs' S (1976, 1988)
- [Statistics legend Ross Ihaka reflects on his revolutionary software](#)
- [The R Language The Good The Bad & The Ugly • John Cook • GOTO 2012](#)

- **No really, what is R?**

- To me,

- R is not a language.

- R has a language.

- R is an environment for doing analytic works and it has a language.

- **R programming language**
  - R language is Open Source
  - R language has excellent tools for data visualization
  - R language has a robust and vibrant online community
  - R language has a powerful package ecosystem

- R is easier to learn for those who do not have computer programming skills.
- R was invented by people who were not trained in Computer Science (CS) and I like this!
- It allows users to start **executing data analyses immediately**



- *"R is more about sketching, and not building. You won't find R at the core of Google's page rank or Facebook's friend suggestion algorithms. Engineers will prototype in R, then hand off the model to be written in Java or Python."- says Michael Driscoll, CEO of Metamarkets.*
  
- *Bow Cowgill, Google "The best thing about R is that it was written by statisticians. The worst thing about R is that it was written by statisticians."*

- The way CS people talk to computers can be a bit strange to us non-CS folks.
- Python was created with traditional CS programming principles in mind
- For instance, is not it more natural to refer to the first element of a five-element set as element 1 and the last as element 5?

Do not say this to CS people; they will not understand.  
For them, the first element is 0 and the fifth is 4.  
Thankfully, R is not like this; the first element of a set in R is referred to as 1, and if there are five elements, the last is 5.

Is not it more suited to the human brain? Absolutely!

- Some differences of R

*assignment operator*

$x = 1$

vs

$x <- 1$

$x = x + 1$  ????

makes absolutely no sense in mathematics.

$x <- x + 1$

- **Some differences of R**

***indexing***

in python

`currentline[3:5]` gives which elements?

3rd and 4th

***in R***

***currentline[3:5] gives ??***

*3,4,5---> makes more sense, it is more sane*

- **Some differences of R**

- Power operator***

- ***In R***

- $3^2 = 9$

- **in python**

- $3**2=9$

- Default functions***

- mean, sqrt etc. available in default R

- Requires additional packages in python

○ Its syntax is concise and easy to read, making it ideal

○ **Example**

- Set all NA elements of x to 0.

- $x[is.na(x)] \leftarrow 0$

- $z \leftarrow \log(x[y > 7])$

○ Fit a linear regression model to w as a function of x, y, and z, including a constant term and all first order interaction terms except xt.

- $model \leftarrow lm(w \sim (x + y + z)^2 - x:z)$

- Example: Simple regression

Growth	Tannin
12	0
10	1
8	2
11	3
6	4
7	5
2	6
3	7
3	8

```
data <- read.table("example.txt", header=T)
```

```
model <- lm (growth ~ tannin)
```

```
summary(model)
```

```
Coefficients:
```

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.7556     1.0408   11.295 9.54e-06 ***
tannin       -1.2167     0.2186   -5.565 0.000846 ***
...
```

```
Residual standard error: 1.693 on 7 degrees of freedom
Multiple R-squared: 0.8157,    Adjusted R-squared:
0.7893
F-statistic: 30.97 on 1 and 7 DF,  p-value: 0.0008461
```

```
# generate and store one million random values  
x <- rnorm(1e6)  
y <- sum(x)
```

Good R style, bad C style

```
# save memory by generating one random value at a time  
s <- 0  
for ( i in 1:1e6 ) s <- s + rnorm(1)
```

Good C style, bad R style

- For loops are inefficient in R, for example



