

Data Science and Analysis with R

By

Dr.Parkavi.A

Assistant Professor

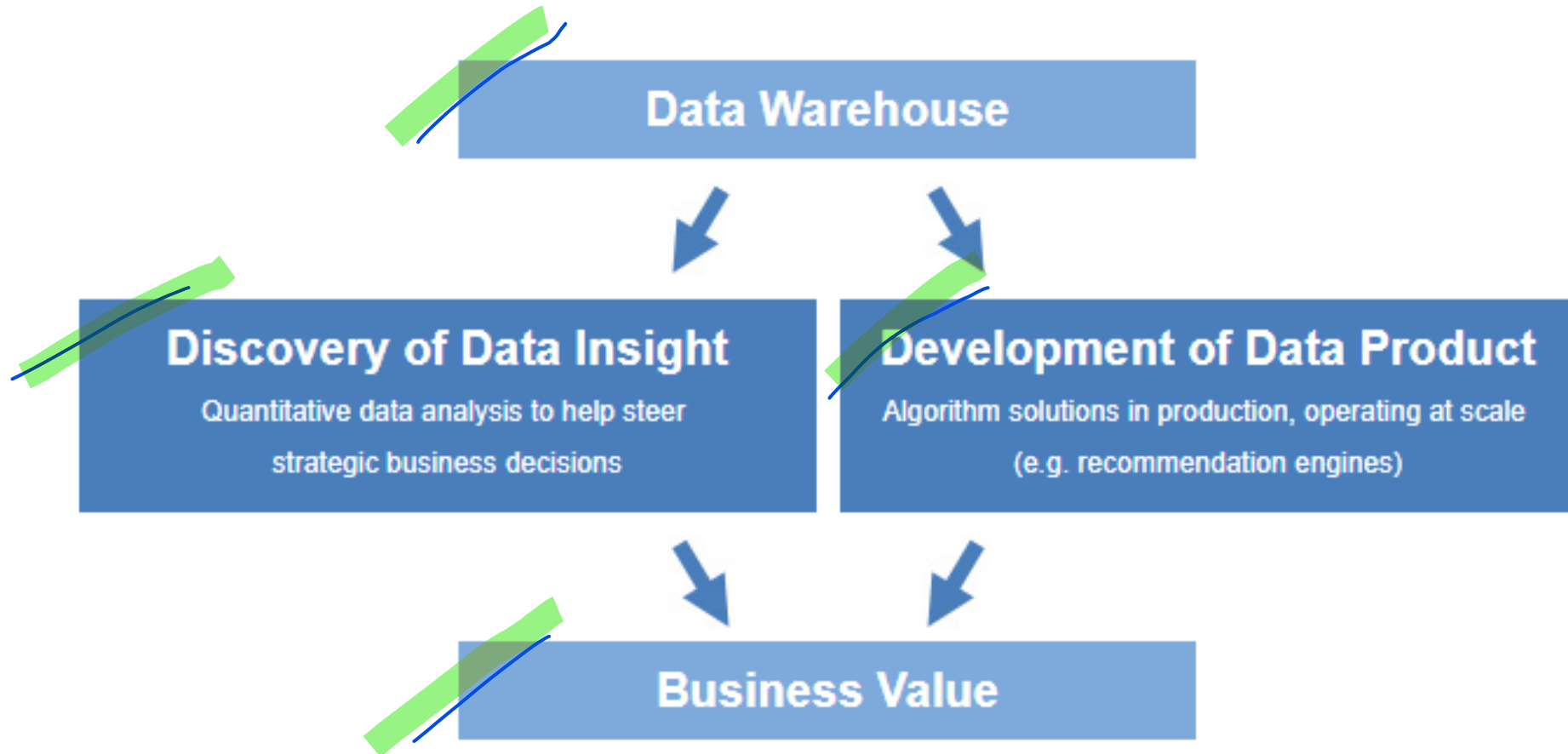
CSE Department

Ramaiah Institute of Tehcnology

Data Science and Analysis

- What is big data?
 - Volume
 - Velocity
 - Variety
- Big data challenges
 - Dealing with data growth
 - Generating insights in a timely manner
 - Recruiting and retaining big data talent
 - Integrating disparate data sources
 - Validating data
 - Securing big data

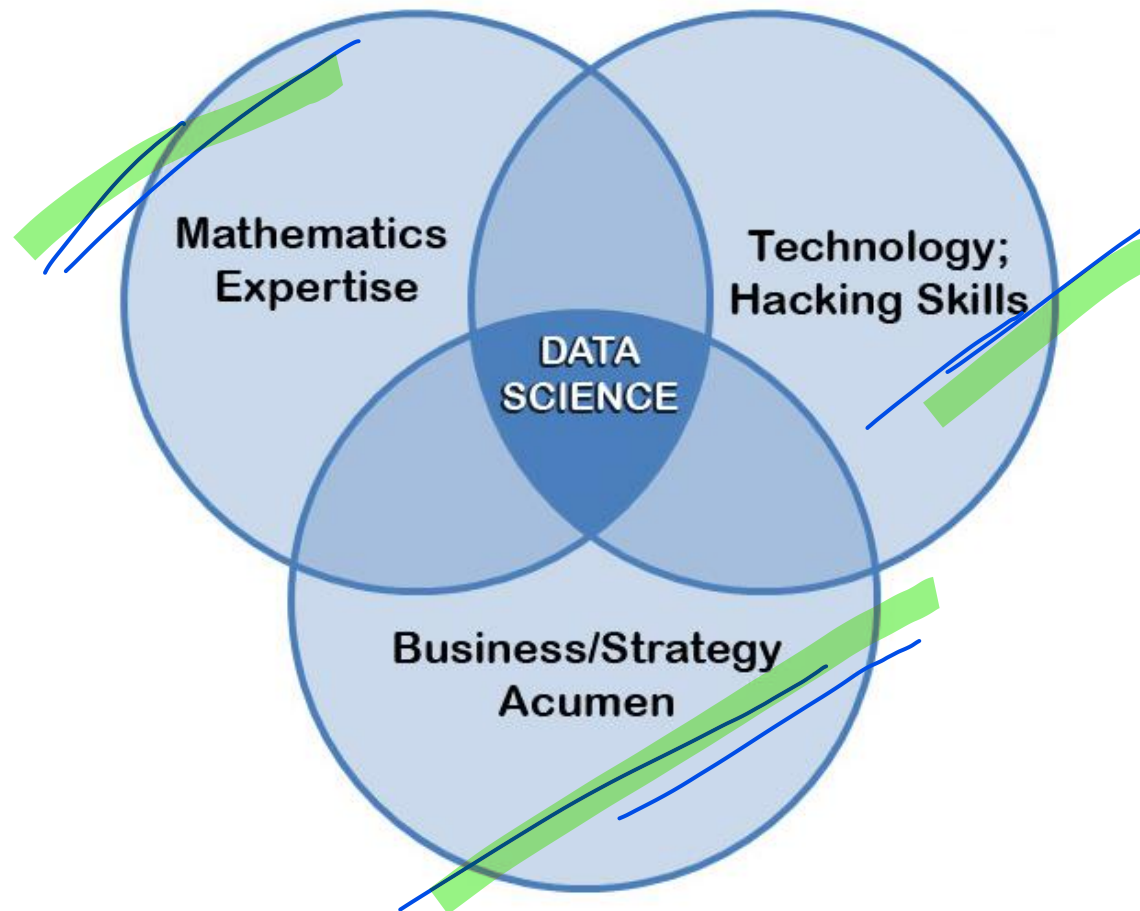
Why Data Science?



Examples: Data Science products

- Amazon's recommendation engines
- Gmail's spam filter
- Computer vision used for self-driving cars

What is data science?



Data Science and Analysis with R

- **Data** manipulation
- Statistical modeling
- Graphics
- Extensibility
- Add-on packages

The R User Interface

- Rstudio

Examples:

To do arithmetic:

```
>1+1
```

```
----
```

To print series of numbers:

```
>100:130
```

```
# Generate a series of numbers from 100 to 130
```

```
series <- 100:130
```

```
# Print the series
```

```
print(series)
```

```
----
```

The R User Interface

- Example:
- Helps to Finish command:

>5+

+

+1

- Examples

>2*3

Objects

What's an object?

- Just a name that you can use to call up stored data

Example:

```
>Mark <- 45
```

```
>Mark +1
```

```
>die <- 1:6
```

```
>die
```

- Working with objects: Vector operation

```
>die-1
```

```
>die/2
```

- To list all objects

```
Ls()
```

BuiltIn- Functions

Examples:

- `round(3.1415)`
- `factorial(3)`
- `mean(1:6)`
- `mean(series)`
- `sample(x = 1:4, size = 2)`
- `sample(x = die, size = 1)`
- `sample(die, 1)`

Writing Your Own Functions

Example

```
roll <- function() {  
  die <- 1:6  
  dice <- sample(die, size =  
    2, replace = TRUE)  
  sum(dice)  
}
```

- Call a function
> roll()

```
Add <- function(a,b)  
{  
  c=a+b  
  print c  
}
```

Add(2,3)

Scripts

- Group of R statements
- Save script
- Run
- Source

Packages

- *ggplot2* : graphics
- To *install package example*:

```
>install.packages("ggplot2")
```

```
>qplot
```

```
>library("ggplot2")
```

```
>qplot
```