```
library(dplyr)

rladies_global %>%
  filter(city == 'Bucharest')
```

DATA ANALYSIS - FROM EXCEL REPORTS TO R NOTEBOOKS

REAL-LIFE USE CASES BROUGHT FROM A DIGITAL ANALYTICS TEAM

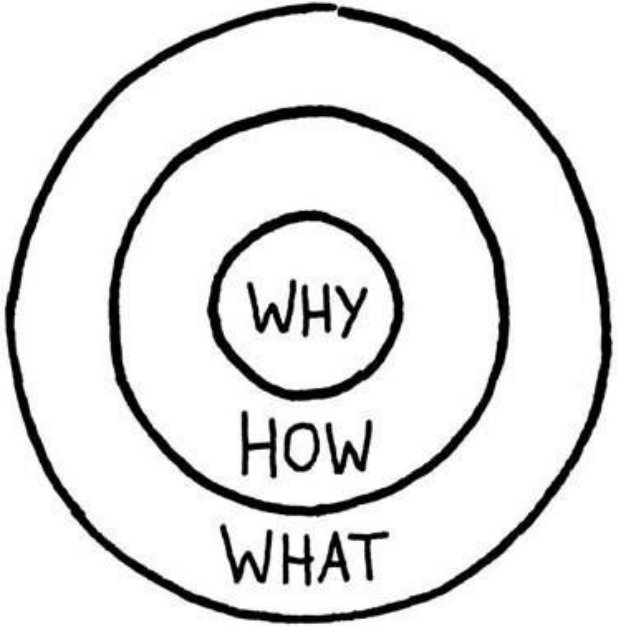
Andrei Deușteanu & Ines Teacă

“

***Andrei, 2Performant Data Expert**, wearing several hats from Data Analyst to Data engineer in his journey of monetizing data value. Indicators of any type, OKRs or KPIs, new methods or algorithms, DW, Big Query, ETL flows, Data Studio, Python or R are some of the tools that is using for supporting his colleagues from business side to deliver value inside the company.*

***Ines, Data analytics specialist,** wearing several hats, from Data Analyst to Product Owner. Her belief is that for any situation there is always a solution! This is why she believes that it’s not enough to have the tools, if you don’t have a great team beside. She uses any available tool to spot the “light”, but she prefers R. :-)*

Agenda



- De ce sa treci de la Excel la R?
- Cum sa faci o analiza?
- Exemple de analiza in R Notebook

Bucharest

1. De ce R Notebooks vs Excel?

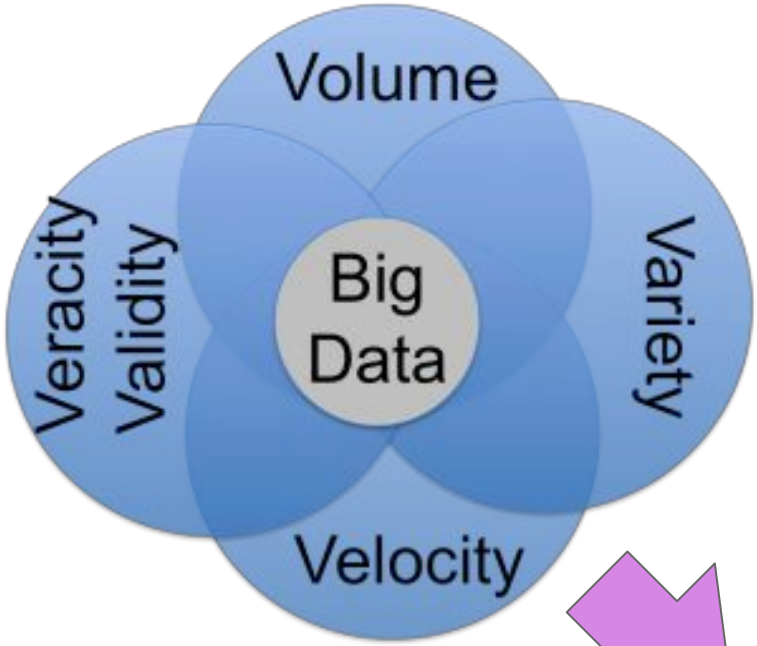
- Care este rolul unui Data Analyst?
- Diferenta intre un raport si o analiza
- Analiza de date “pe coate”

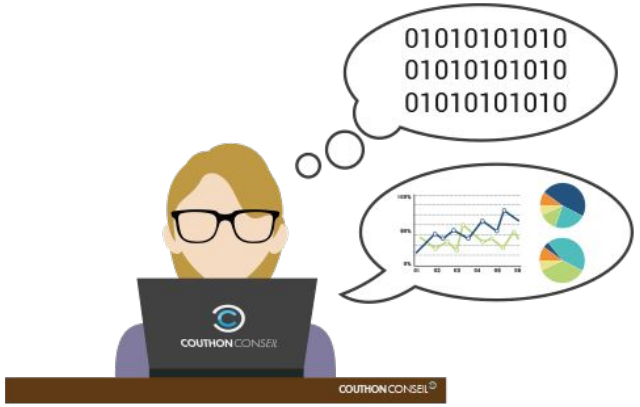
Bucharest

Rolul unui Data Analyst

=

Business Person


Volume
Veracity
Validity
Big
Data
Variety
Velocity


01010101010
01010101010
01010101010

Bucharest

Turn the 4 Vs of Big Data into the single V of Business - Value

- Volume
  - RAM is the limit, not Excel
  - Daca te superi folosesti procesare in paralel
- Variety
  - TXT, CSV, JSON, SQL etc. -> I can handle it!
  - Finance, Marketing, Sales, Operations, Commercial - it’s just data to me
- Veracity
  - Definitii de business + SQL
  - Problemele datelor pe care le colectam
- Velocity
  - Scrii cod SQL mai repede decat seful tau poate sa pronunte Indicatori Cheie de Performanta

Abilitati cheie ale unui analist

1. Procesarea datelor
   a. Citire
   b. Curatare
2. Explorarea datelor
   a. Inteles de business
   b. Agregari
   c. Vizualizari
3. Extragerea concluziilor relevante
4. Sintetizare

Data

MAGIC

Insight

Analiza vs Raport

Bucharest

Raport -> “Vreau sa explorez aceste date mai departe”

Analiza -> “Am nevoie de cateva concluzii in baza carora sa actionez”

https://2performant.com/ecommerce-statistics/

Zona de Insights = Analiza

Zona de Intrebari cu embedded Data Studio = Raport

De multe ori oamenii cer un raport cand de fapt au nevoie de analiza.

E treaba analistului sa-si dea seama ce are omul nevoie de fapt.

5 min de discutie inainte de a te apuca de lucru te pot salva de multe frustrari ulterioare

Analiza de date “pe coate”

Bucharest

“Stim ca cele mai multe achizitii vin din paginile de produs. Care sunt top tipuri de pagini din care oamenii ajung in paginile de produs? Dupa ce le aflam o sa lucram sa le mai optimizam ca UX”

Google Analytics -> CSV -> Procesare de text in R -> Pivot & Pivot Chart in Excel -> Grafice in PowePoint -> Share pe FileServer -> Concluziile pe mail -> Done! 🙂

Probleme:

1. Nu are acces pe File Server -> obtine de la IT
2. Are Mac & alta versiune de Office, graficele nu se vad bine
3. Am gresit ceva la extragerea datelor => **totul a trebuit refacut** 😭

Bucharest

2. Cum sa faci o analiza?

- Problema cu analiza “pe coate”
- O varianta mai buna -> R Notebooks
- Cum sa structurezi o analiza

Probleme cu Analiza de date “pe coate”

- Acces la date (drepturi etc.)
- Lipsa de portabilitate (MacOS vs Windows, Versiuni de Office)
- Lipsa de reproductibilitate (intre colegi)
- Repetitivitate (editarea graficelor ca sa arate mai bine)
- Daca ai gresit ceva trebuie sa reiei tot procesul

O varianta mai buna - R Notebooks

Bucharest

- Extragerea datelor, procesarea, analiza, concluziile sunt toate in acelasi sistem
- Output as HTML => Orice browser il deschide
- Scrii codul 1 data si rulezi de cate ori doresti
- Toti pasii sunt scrisi direct in cod -> poti share-ui cu alti analisti

Structura unei analize de date

Inspirat din CRISP-DM

1. Intelegerea Business-ului
   a. Intrebare / Ipoteza
   b. Definit suficient de concret pentru SQL
2. Intelegerea Datelor
   a. Ce date am nevoie?
   b. Ce date sunt dipsonibile?
   c. Unde sunt datele?
3. Prepararea Datelor
   a. Aduc datele in mediul de lucru
   b. Le combin in formatul in care am nevoie
4. Modelare / Explorare
   a. Vizualizari
   b. Agregari
5. Evaluare / Sintetizare
   a. Ma uit la Vizualizari / Agregari
   b. Aleg ce e mai relevant
   c. Formatez graficele & textele sa arate mai bine
   d. Concluzii
6. Deployment / Sharing
   a. Fisierul HTML cu analiza

Bucharest

Bucharest

3. Exemple de analize in R Notebooks

- Studiu de caz 2Performant - Advertiserii care accepta PPC au rezultate mai bune?
- Studiu de caz eCommerce - Un raport rapid in Excel (demo)

Bucharest

Advertiserii care accepta PPC au rezultate mai bune?

Studiu de caz cu date din platforma 2Performant

Bucharest

Inca un raport rapid in Excel (demo)

Studiu de caz eCommerce

Un workflow rapid si simplist

Bucharest

openxlsx

Export results from R to Excel

RMySQL

Connect to db & bring the data in R

tidyverse / dplyr

Prepare easily the data in R

De unde aduc datele?

Bucharest


```
install.packages("RMySQL")
library(RMySQL)
```

mydb = dbConnect(MySQL(), user='user', password='password', dbname='database_name', host='host')

Create a database connection object.

dbListTables(mydb)

This will return a list of the tables in our connection.

dbListFields(mydb, 'some_table')

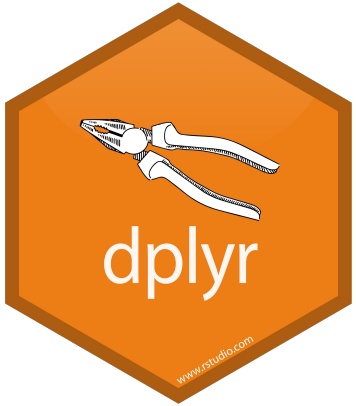
This will return a list of the fields in some_table.

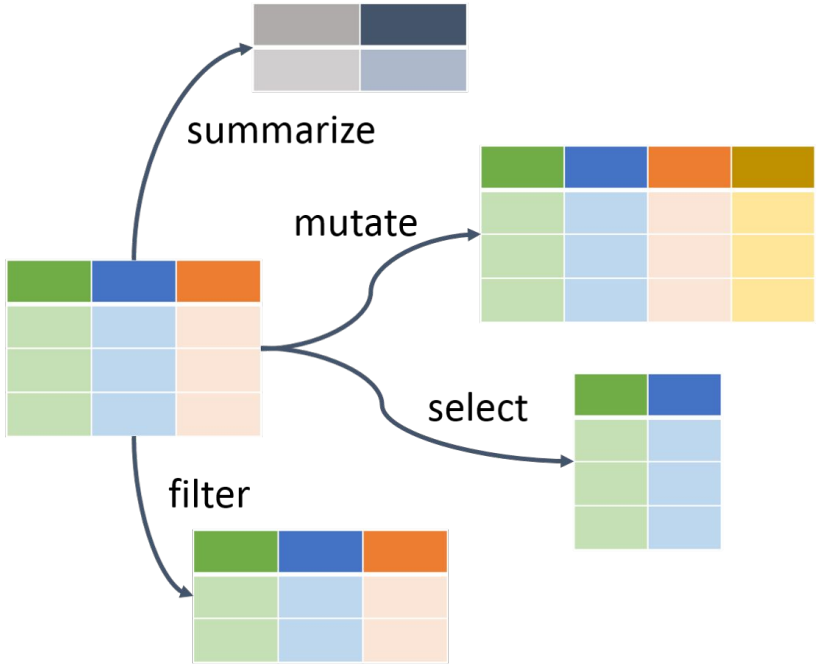
dbReadTable(mydb, 'table')

Read table from database

Pot sa le mai “aranjez” un pic?

Bucharest






```
install.packages("dplyr")
library(dplyr)
```

Si acum.. raportul rapid pe coate

Bucharest


```
install.packages("openxlsx")
library(openxlsx)
```

wb <- createWorkbook()

Create an Excel workbook.

addWorksheet(wb, "table")

writeData(wb, sheet = "export_table", x = table, headerStyle = header_style)

This will create and write a sheet in the workbook we created.

saveWorkbook(wb = wb, file = file, overwrite = T)

This will save the result as xlsx file.

url: https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf

Cele 5 idei pentru automatizarea mea cu R

Bucharest

1) Totul se intampla rapid - si tu trebuie sa te misti rapid, dar atentie! incearca sa intelegi corect cerinta primita
2) Incepe prin a construi pas cu pas procesul avand in vedere trecere intr-un setup de reproducere ulterioara a analizei
3) Renunta la coloanele cu SUMIF, COUNTIF si VLOOKUP ce dureaza o ora pentru secundele din tidyverse/dplyr cu summarize
4) Folseste parametri. De exemplu, data de sistem
   E.g. format((Sys.Date()-1),"%Y%m%d");
5) Foloseste functia **paste**

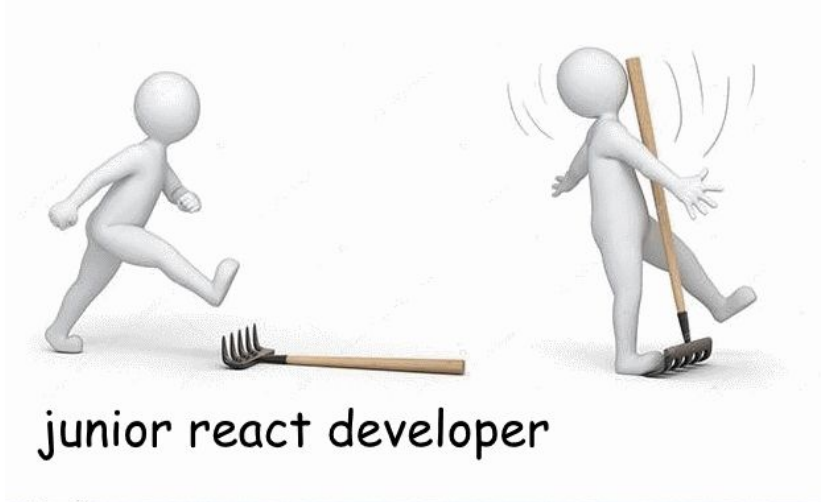
```
E.g.  a<-'Raport'
      b<-'.xlsx'
      n1<-paste(a,b, sep="")
```

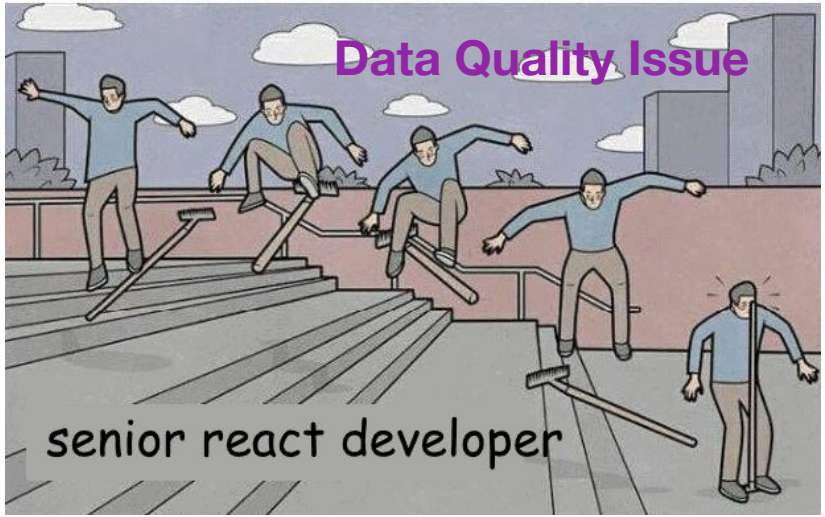
Ines Teaca

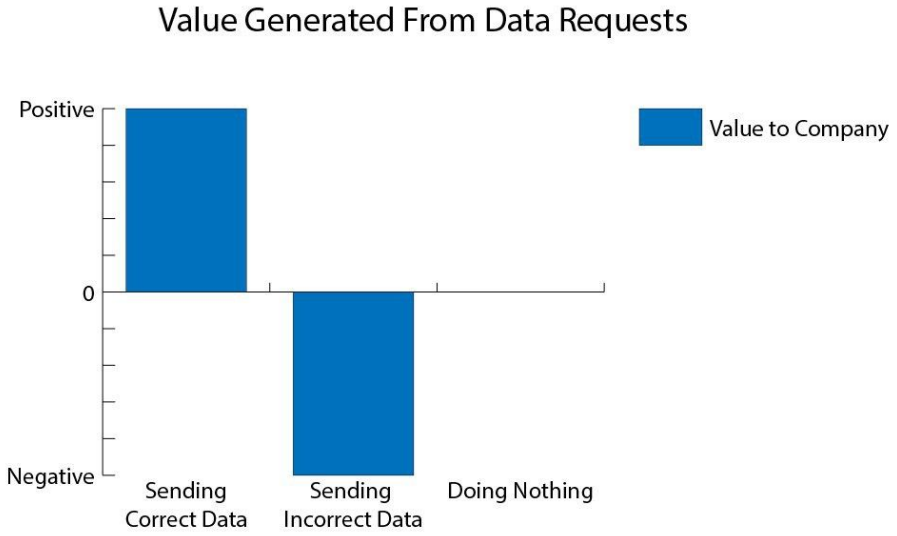
Concluzii

- Analiza sau Raport
  - Explorare sau Concluzii?

- Excel sau Notebook?
  - Volum de date >1 GB?
  - Cat de complexa e pregatirea datelor?
  - Cat de multe & cat de complexe sunt vizualizarile / agregarile?
  - Cat de importanta este reproductibilitatea?







Rolul Analistului - Tranforma cei 4 V ai Big Data intr-un singur V - Valoare

- Cum faci asta este treaba ta
- **Important este sa-ti faci munca mai usoara & sa oferi date corecte**

Bucharest

THANK YOU
MULȚUMIM

next meetup → Tue,10 March 7.00PM

V7 Startup Studio HUB

#eachforequal #IWD2020

www.rladiesbucharest.org