

# Data driven decisions - Day 1

*Understanding data and it's usage*

12. – 14.5.2021



With the support of the  
Erasmus+ Programme  
of the European Union

1

## Agenda for the training

### Wednesday

- Lecture

### Thursday

- Video, no live session
- Working with data in Excel
- <https://youtu.be/4l8GW3i3dsl>

### Friday

- Guest lecturer from an AI company
- Closing thoughts

Data Driven Decisions

2

## Agenda for today

### Lecture

- Where's data coming from
- How data is stored
- What can we do with data

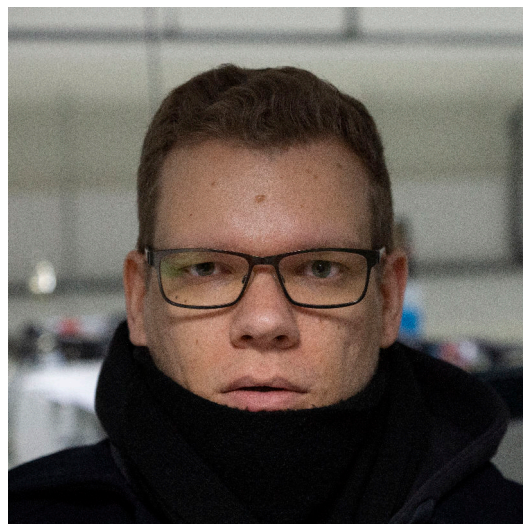
Data Driven Decisions



3

## Juho Salli

- A teacher-researcher
- A father and a husband
- Background in software development and data engineering
- Likes to run (a lot)



Data Driven Decisions



4

# Generating data

Data Driven Decisions



5

## Data sources

- Mobile phones
- Televisions
- Internet browsers
- Sensors in public places
- Cars
- Sport watches
- Games
- PowerPoint slides
- Processes
- Grading students
- Money transactions
- Hospital visits
- Emails
- Books
- ...basically anything that can be written down

Data Driven Decisions



6

# Storing the data

Data Driven Decisions

 TALENTJOURNEY

7

## Data type

- Text
- Pictures
- Sound
- Video
- Executable file



Data Driven Decisions

 TALENTJOURNEY

8

## Storage type

- For structured data
  - Relational database
- For semistuctured data
  - NoSQL database
- Data with no structure
  - Maybe some data lake solution



Data Driven Decisions

 TALENTJOURNEY

9

## Structured data

- Tabular data
- Rows
- Columns
- Possibly some relations between different data points
- The structure doesn't change



Data Driven Decisions

 TALENTJOURNEY

10

## Structured data

Id	Firstname	Lastname
1	John	Doe
2	Jane	Doe
3	Mary	Poppins

Id	Person_id	Eye color
1	2	Green
2	3	Gray

Data Driven Decisions



11

## Semistructured data

- Data might have some structure but it's prone to changes

```
{
  "people": [
    {
      "id": 1,
      "firstname": "John",
      "lastname": "Doe"
    },
    {
      "id": 2,
      "name": "Jane Doe",
      "eye_color": "Green"
    },
    {
      "id": 3,
      "firstname": "Mary",
      "lastname": "Poppins",
      "eye_color": "#808080"
    }
  ]
}
```

Data Driven Decisions



12

# Working with the data

Data Driven Decisions



13

## Data preprocessing

- Data needs to be in such format that it can be used with analytical tools
- Text to number representations
- Numbers to binary
- Combining multiple sources -> harmonization

### Encoding

- "Positive" -> 1
- "Negative" -> 0
- "January" -> 0
- "May" -> 4

Data Driven Decisions



14

## Domain knowledge

- The key factor for successful insights
- Team effort
- Especially when combining different data sources
- Might require a multidisciplinary team



Data Driven Decisions

 TALENT JOURNEY

15

## Visualisation

- A good first step is just to eyeball your data
- Humans are visual beings so we spot visual differences and outliers easily

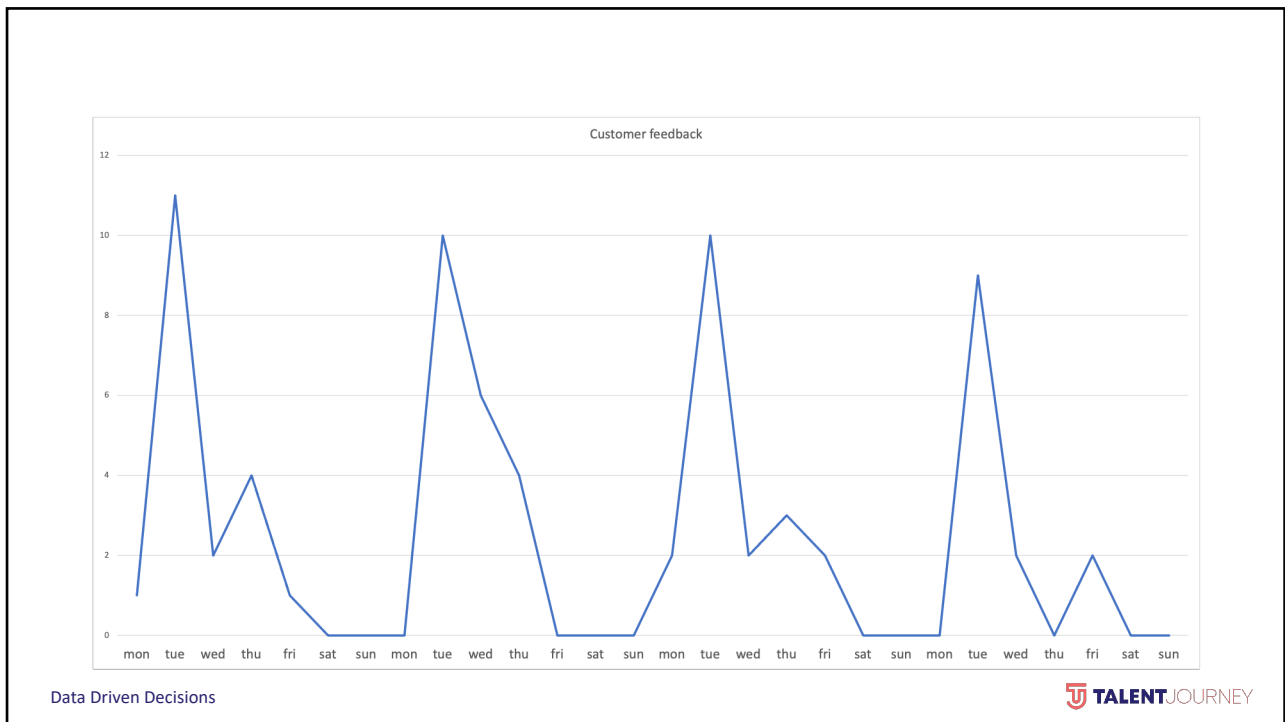


Data Driven Decisions

 TALENT JOURNEY

16





17



18

## Tooling

- Start with the simplest tools and methods
- Then work your way up
- Raw data -> spreadsheets -> visualisation -> something else



Data Driven Decisions

 TALENT JOURNEY

19

## Tooling

- Linear regression for simple predictions
- Natural Language Processing (NLP) tools for working with natural language
- Decision trees, neural networks
- Different tools for anomaly detection
- Process mining
- Evolutionary algorithms



Data Driven Decisions

 TALENT JOURNEY

20

## Natural Language Processing

- Finding value in natural (spoken) languages
- Sentiment analysis
- Text to speech
- Machine translation
- A huge field with lots of different categories and applications



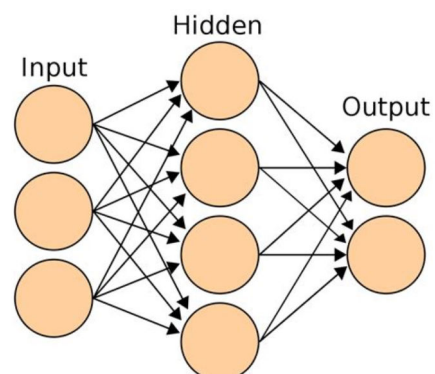
Data Driven Decisions

 TALENT JOURNEY

21

## Neural network

- Draws inspiration from the human brain
- Basically just trial and error until the output is good enough



Data Driven Decisions

 TALENT JOURNEY

22

## Artificial Intelligence

- Works when the problem is narrow
- Works when humans have defined the problem correctly
- Won't take our jobs, but can work with us
- Won't take over the world any time soon (if ever)



Data Driven Decisions

 TALENT JOURNEY

23

## What data can do for us

- Improve existing products
- Improve existing processes
- New products
- New innovations
- Game changers



Data Driven Decisions

 TALENT JOURNEY

24



# Thank you!

See you Friday!

Data Driven Decisions (pictures from unsplash.com or by author)



With the support of the  
Erasmus+ Programme  
of the European Union