Data Analytics

Agenda

1. Data Analytics vs Business Analytics
2. Coursera Data Analysis Course
3. Python Packages

Data Analytics vs Business Analytics

Business Case Studies

- are real-world experiences of how business people and companies succeed or fail
- examine events that have already happened

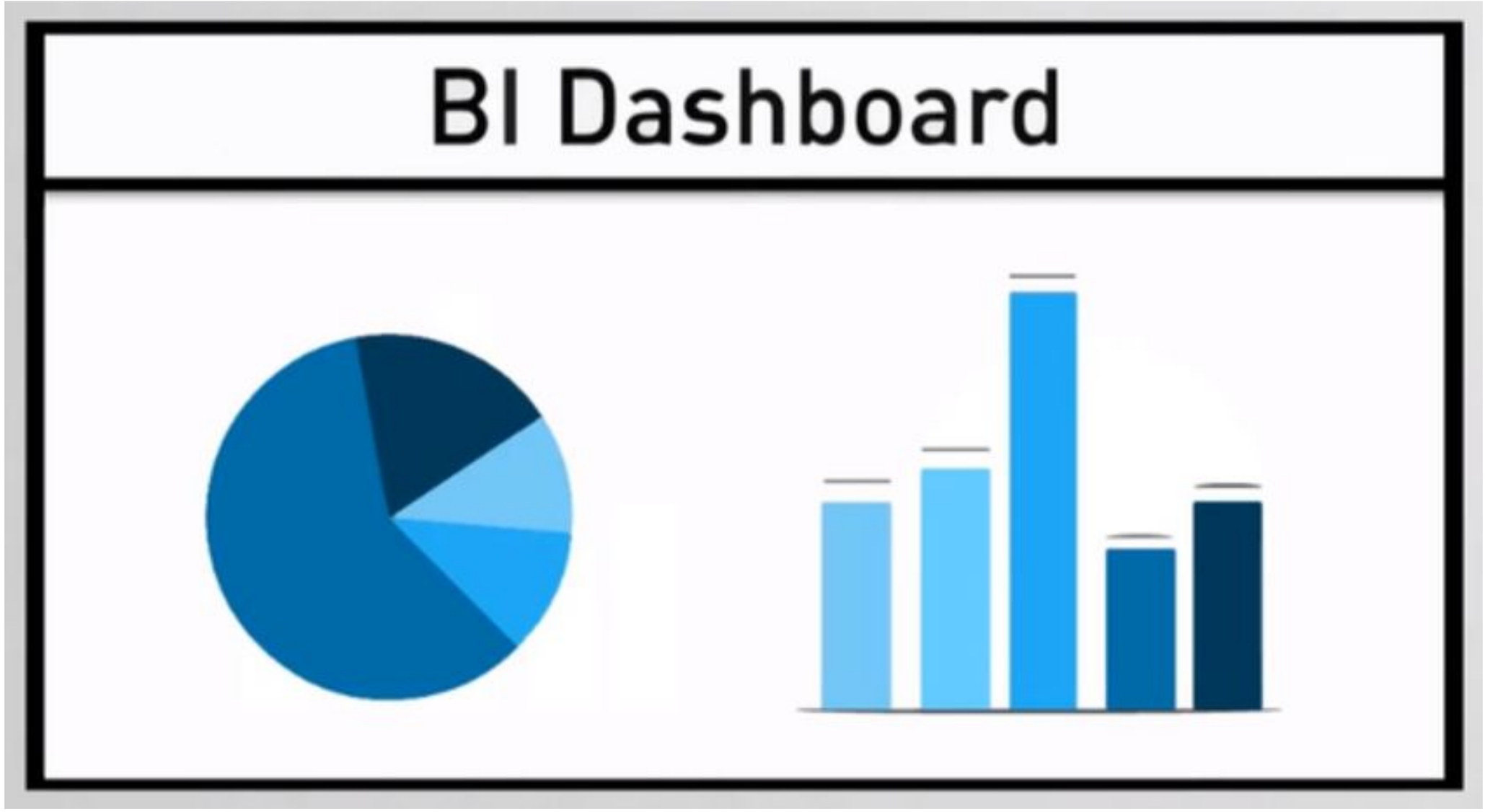
- We do not need a dataset to learn from **business cases**
- We could learn from them and attempt to prevent making a similar mistake in future

Business Case Studies: Example

Ryanair: `cost-friendly budget airline that sells tickets so cheap`

- travel to less busy airports
  - far from the city
  - outside business hours
- charges for almost every small addition
- tries to keep planes for small times on airfields to save on rent
- operates only one type of aircraft to speed out ground crew processes

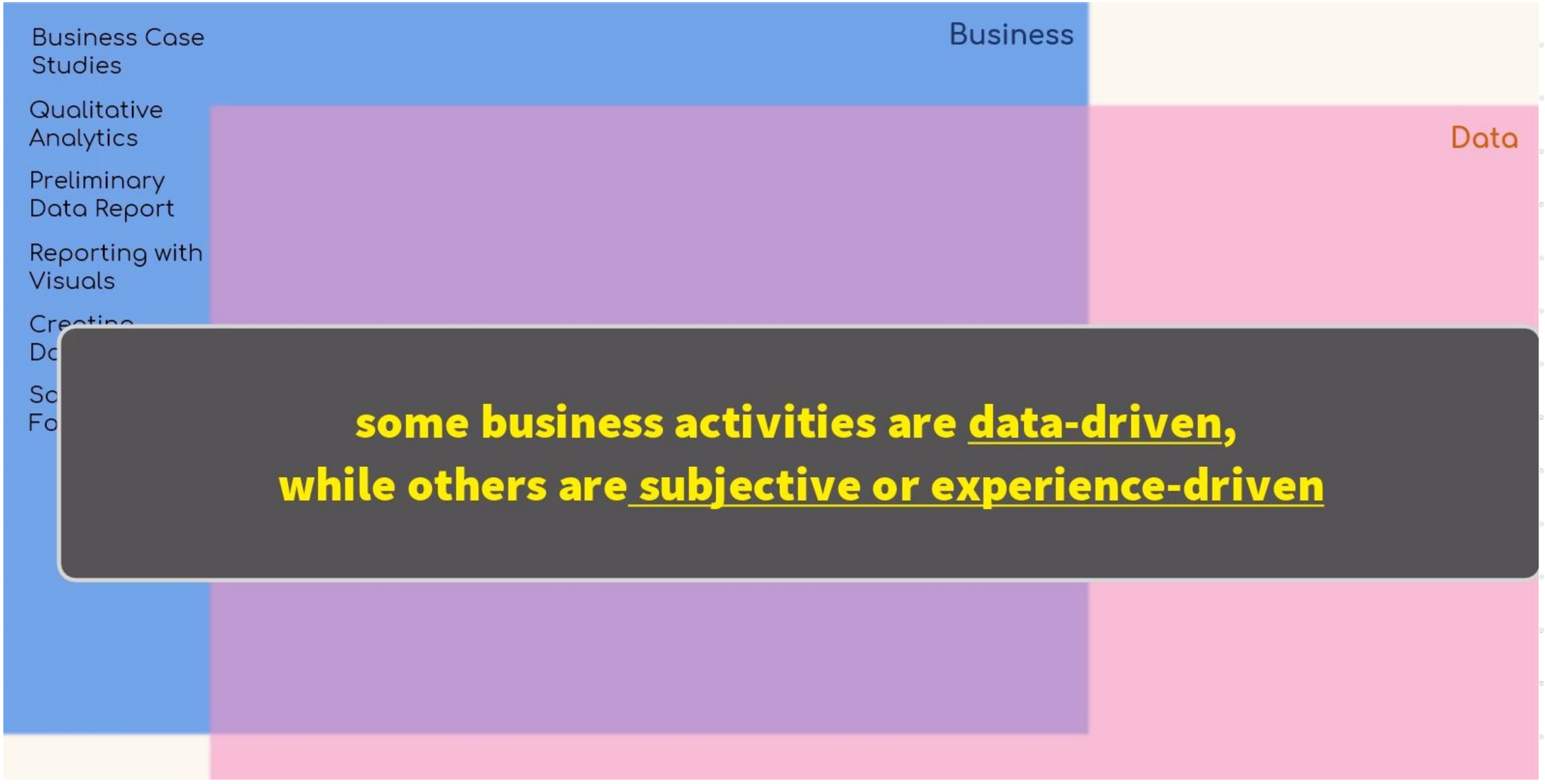
Visuals & Dashboard: Example



Business


Business Case Studies
Qualitative Analytics
Preliminary Data Report
Reporting with Visuals
Creating Dashboards
Sales Forecasting
Business

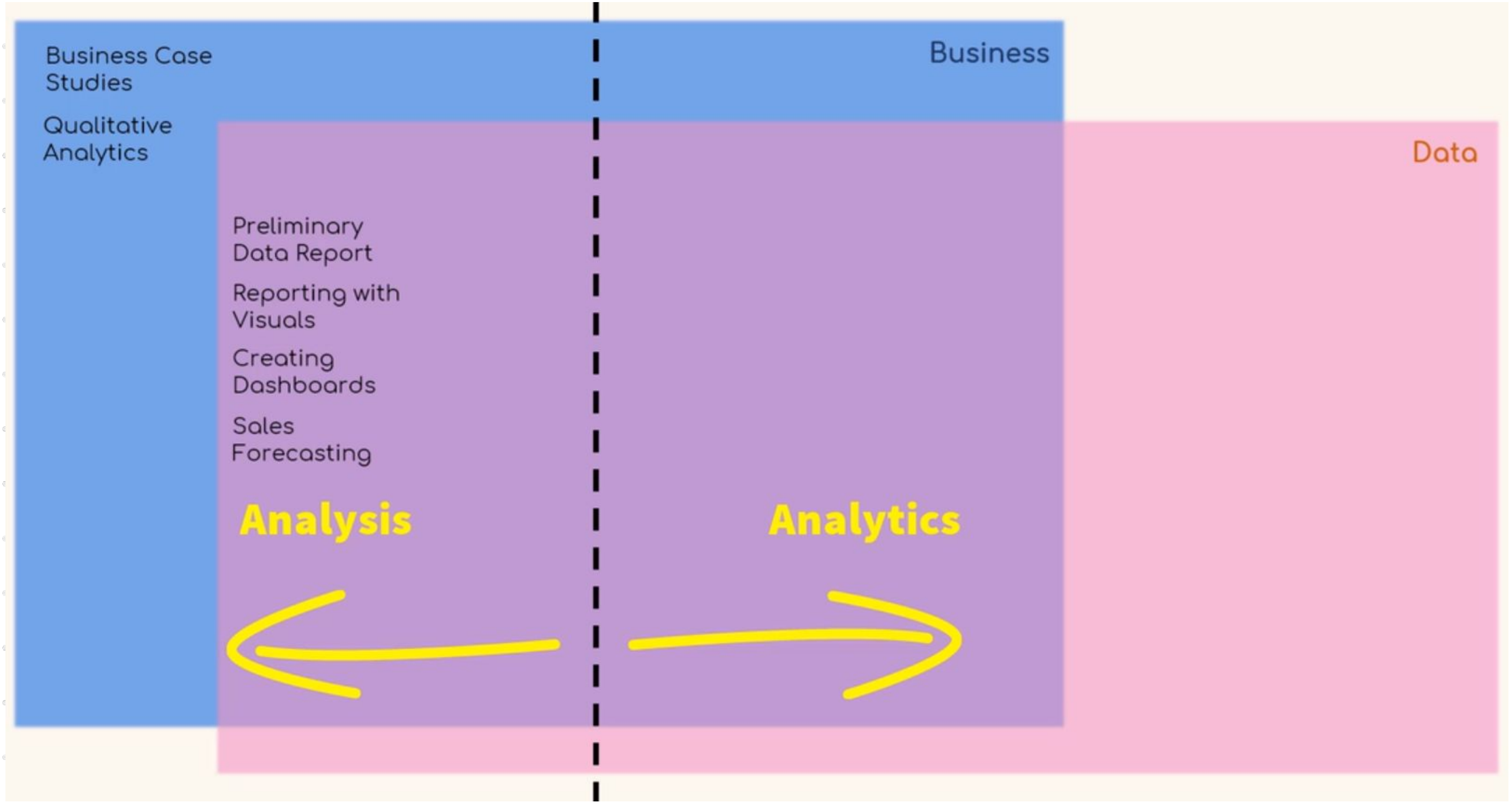
Data?


Business Case Studies
Qualitative Analytics
Preliminary Data Report
Reporting with Visuals
Business
Data
some business activities are data-driven,
while others are subjective or experience-driven

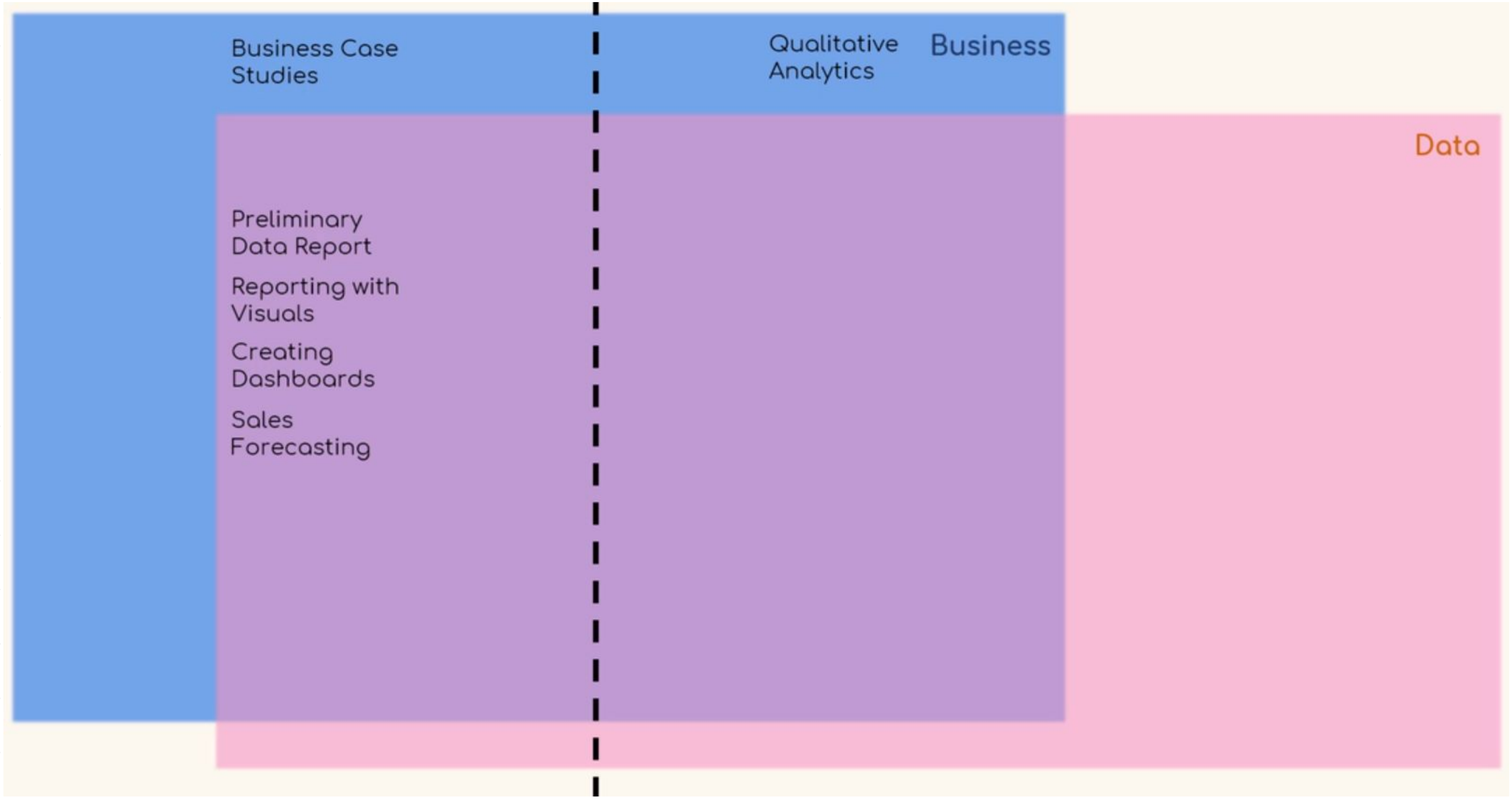
Data


Business Case Studies
Business
Qualitative Analytics
Data
Preliminary Data Report
Reporting with Visuals
Creating Dashboards
Sales Forecasting

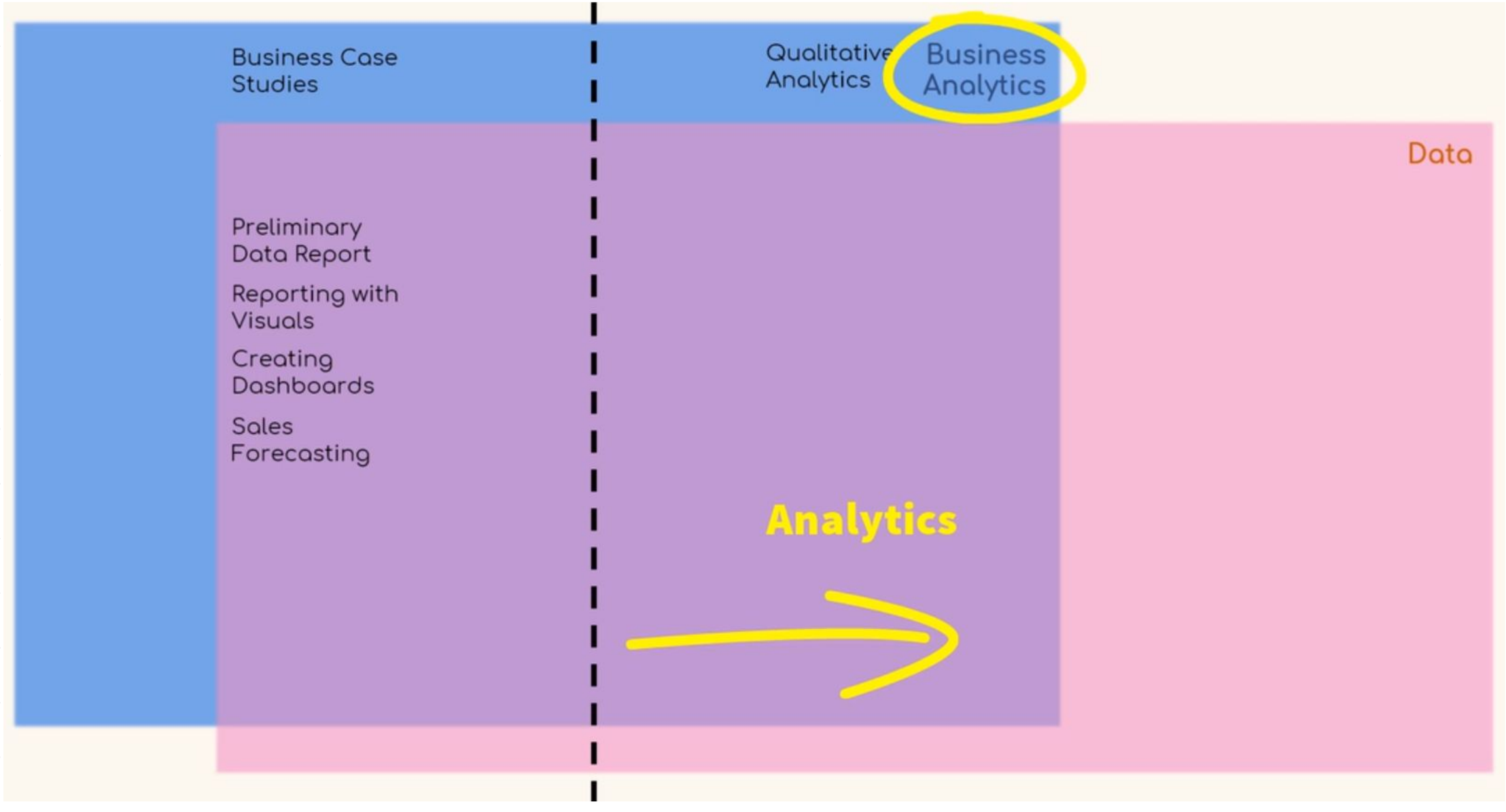
Past vs Future?



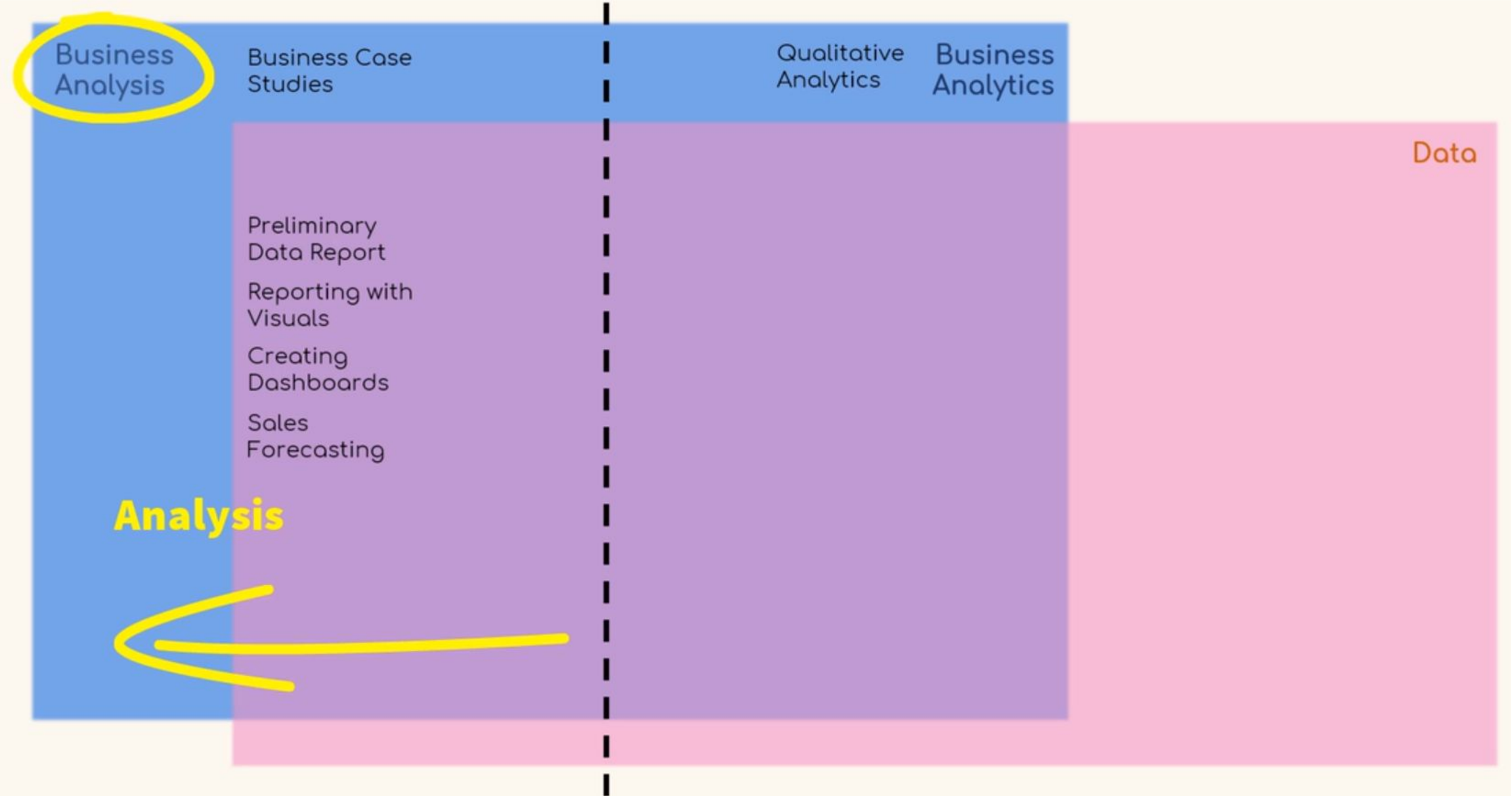
Past vs Future



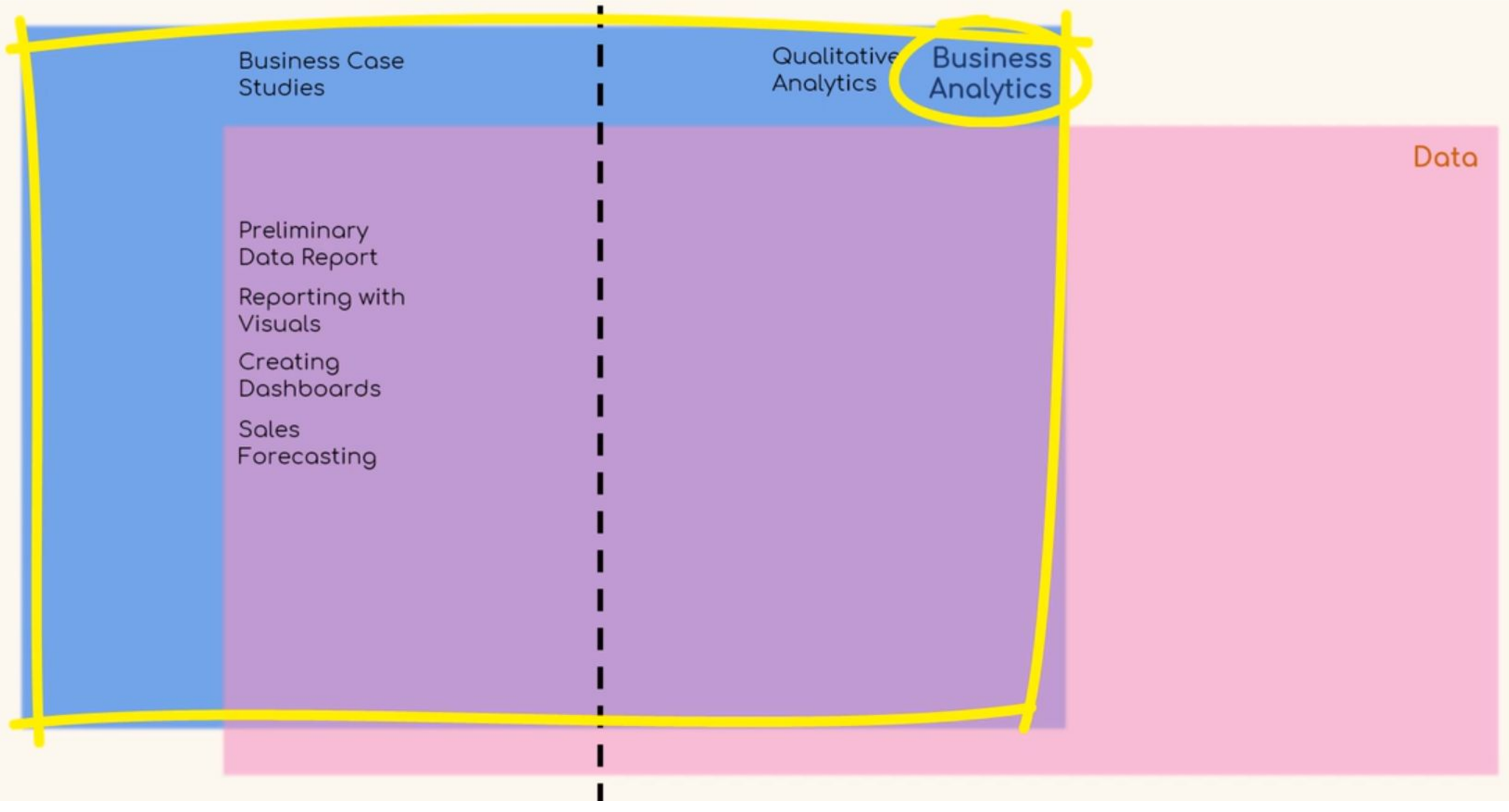
Business Analytics



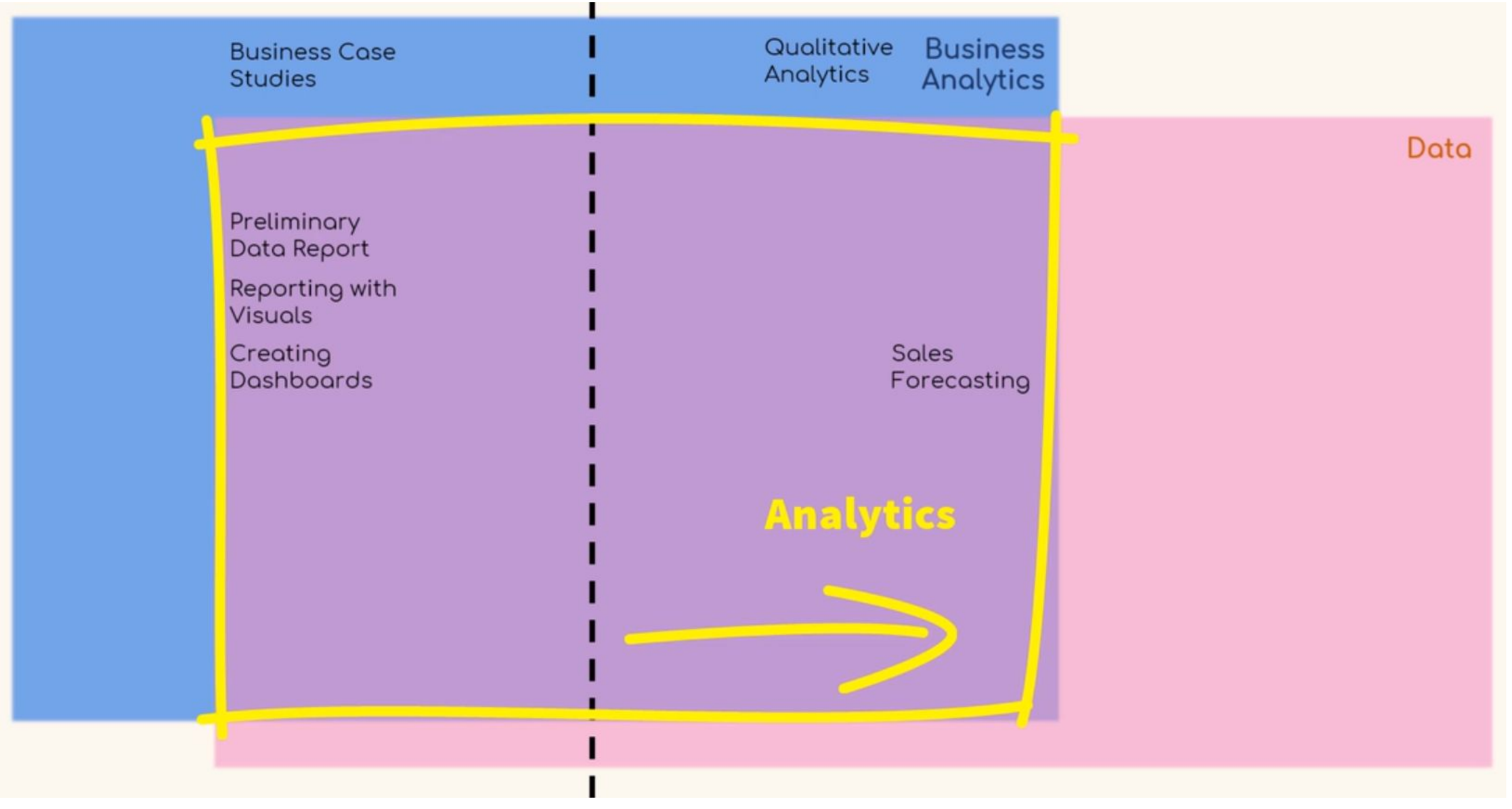
Business Analysis



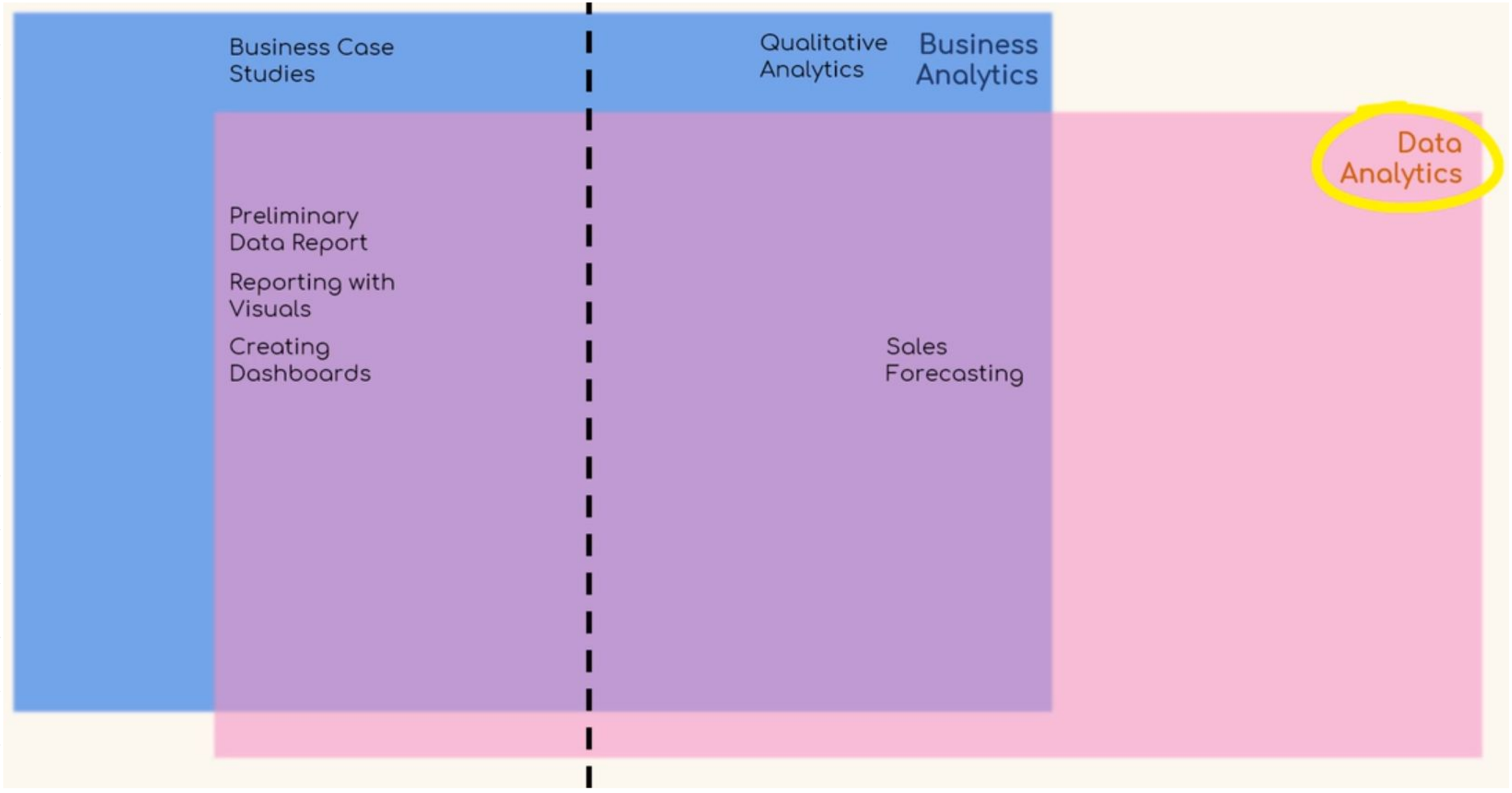
Business Analytics += Business Analysis



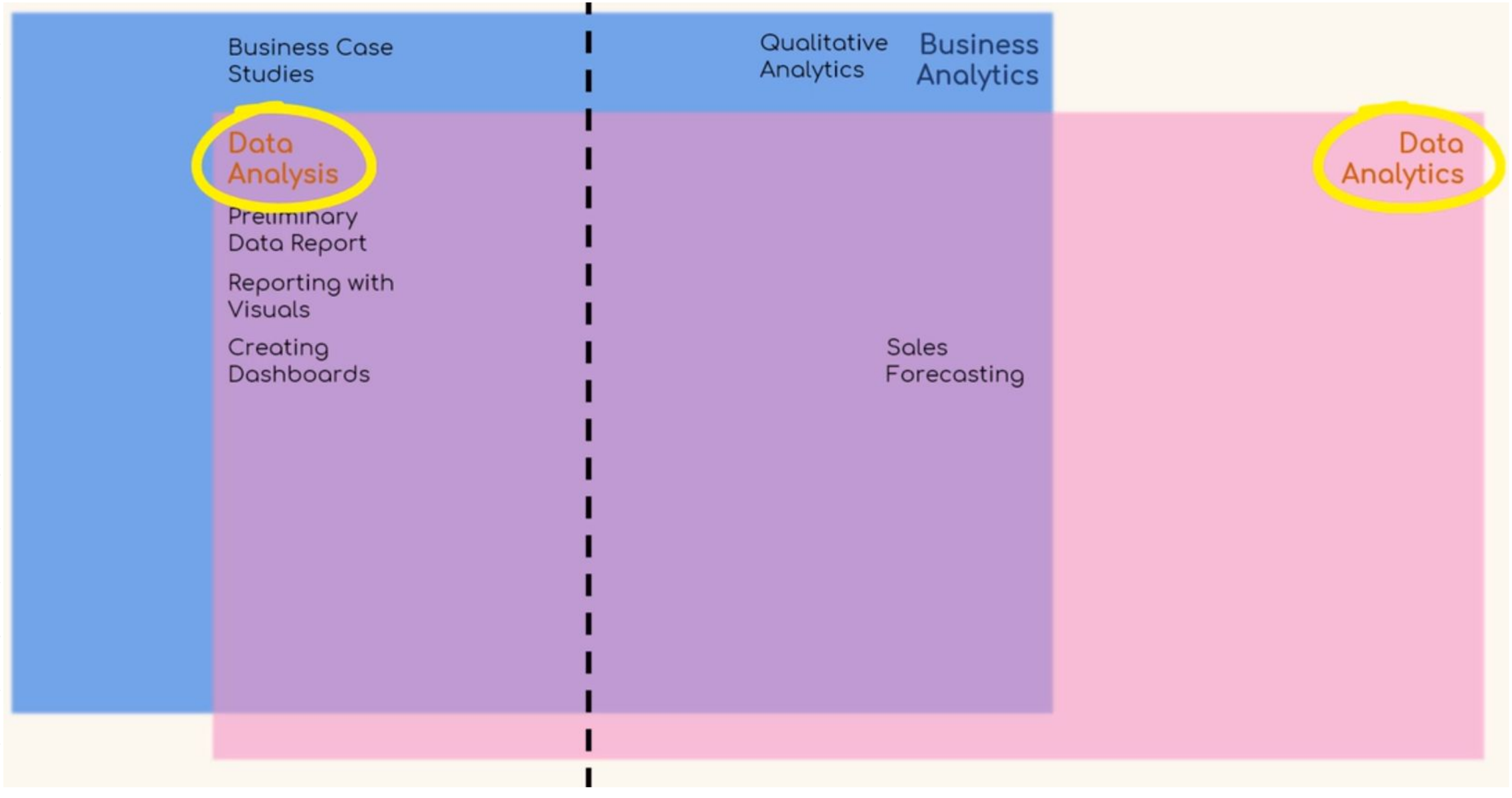
Data Analytics?



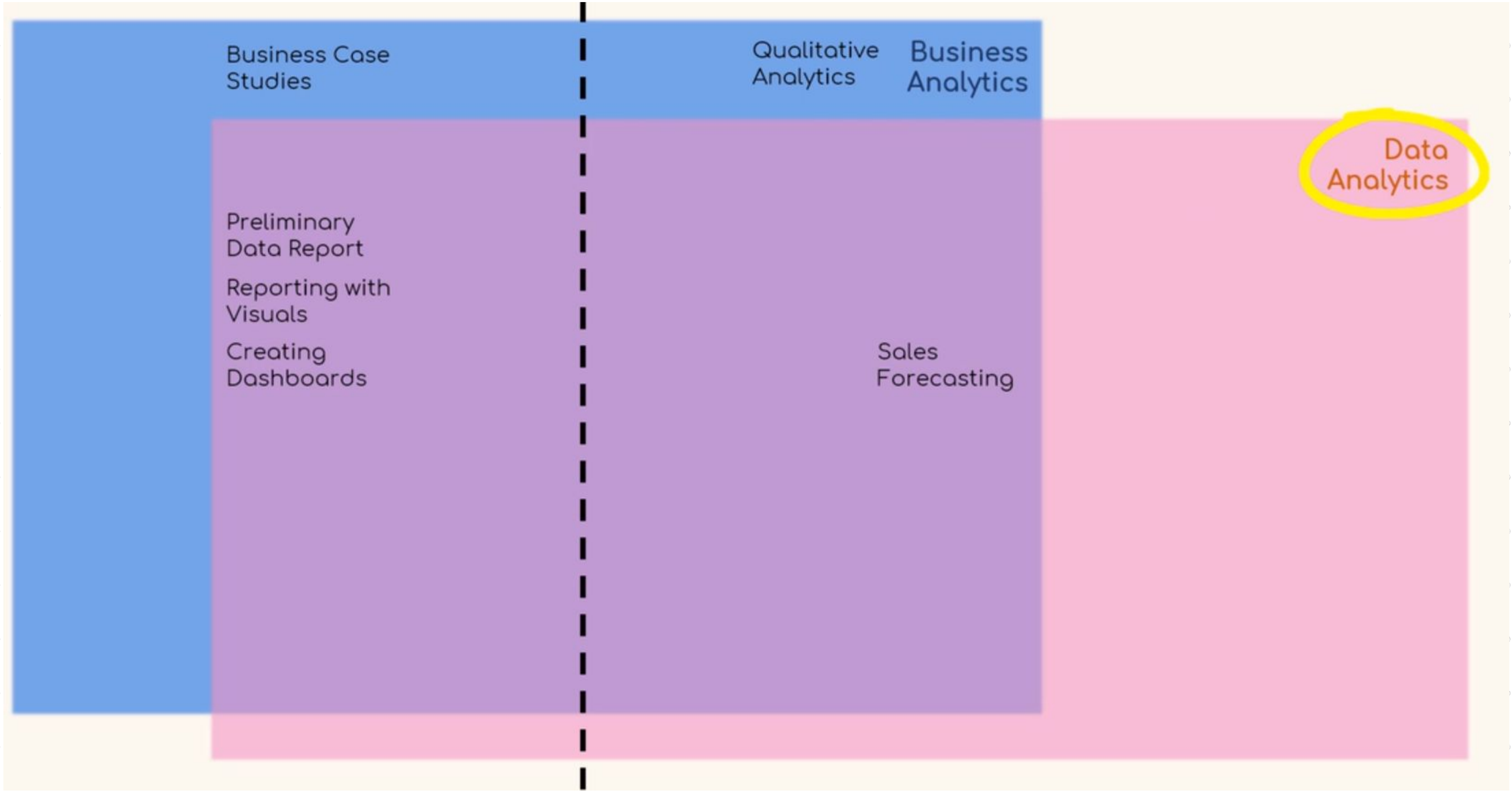
Data Analytics



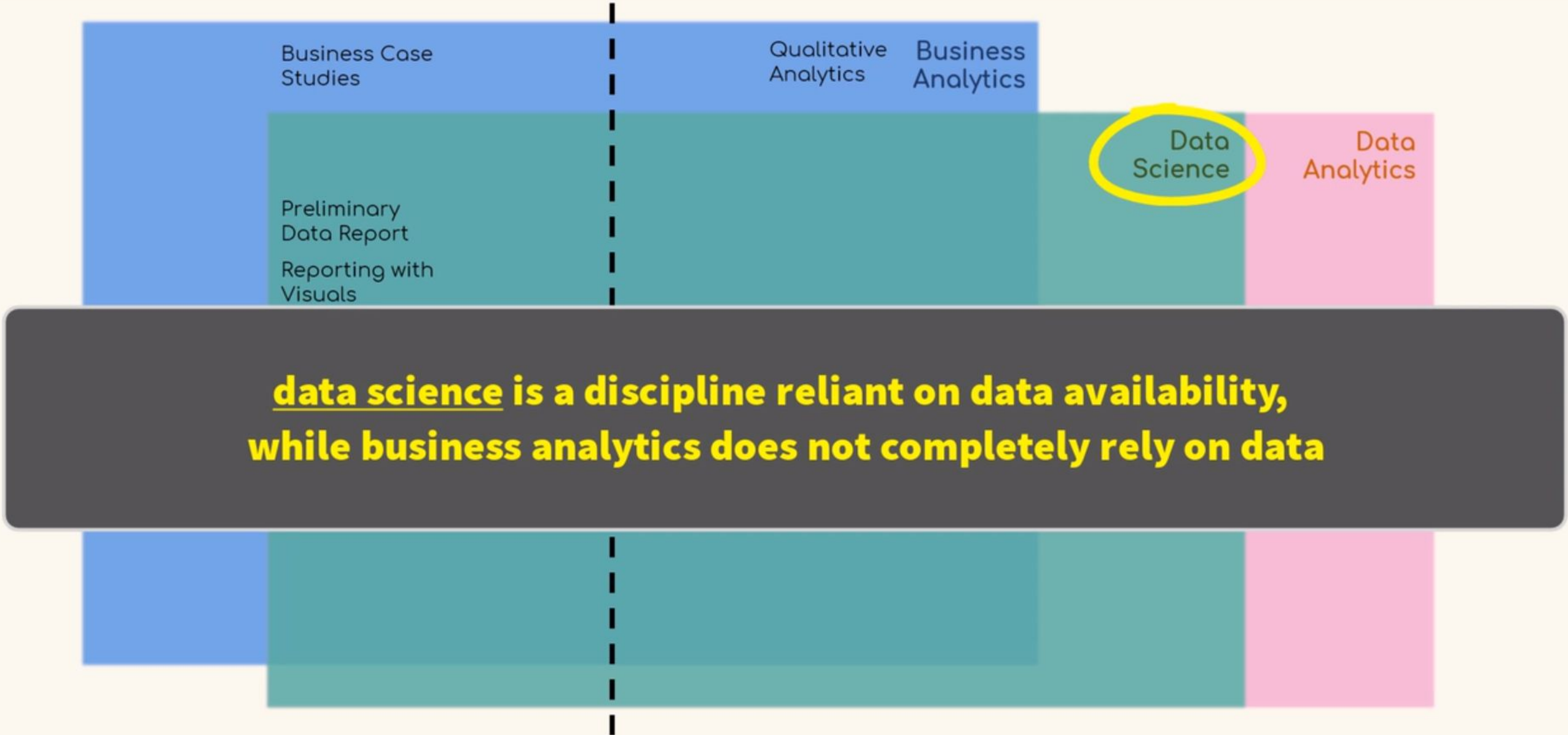
Data Analysis



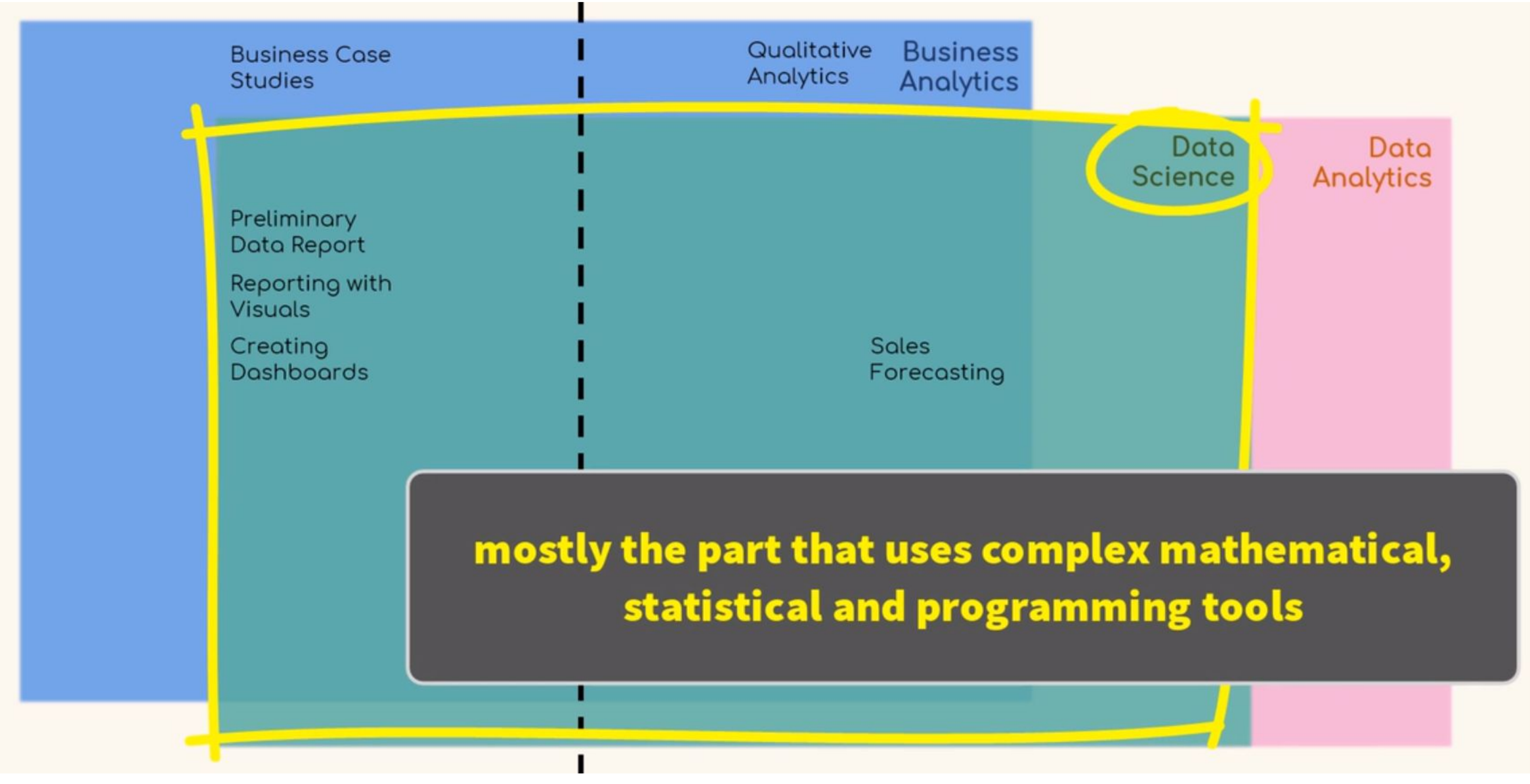
Data Analytics += Data Analysis



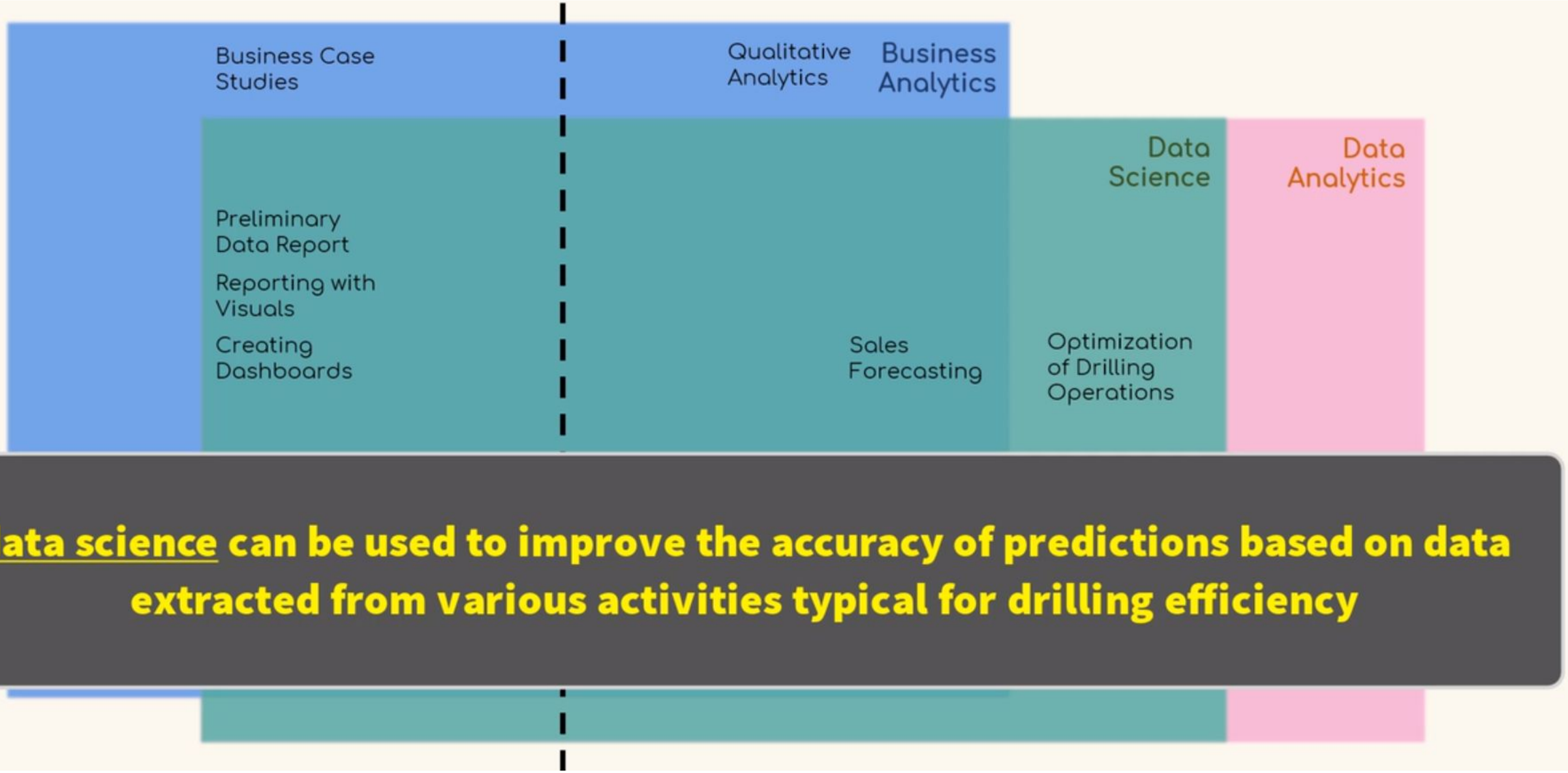
Data Science - 1



Data Science - 2



Drilling Operations Optimization



Drilling Optimization - Course

- 1 day
- Intro Level
- Up to $900

15th October 2018 – Cairo, Egypt

Held in conjunction with the SPE Workshop: Drilling and Optimisation, 14-17 October 2018.

Description

In today's world, traditional methods of drilling oil wells don't work as much anymore. Yesterday's practices are being superseded by a universal trend towards the extensive use of historical and real-time data to understand, learn and predict all well intervention operations. This course explores the impact of data analytics on well operations. Drawn from the presenter's extensive experience in data analysis, it examines, in easily understandable terms, today's data management processes targeting process improvement. The course introduces issues related to:

- Basic concepts of data management with emphasis in drilling operations
- Methods of using data as a basis for improving and optimizing process control
- Achieving a common understanding of the issues involved among information technology personnel and field engineers
- A roadmap for the implementation of a drilling process improvement system
- Business Intelligence as the ultimate goal of data management processes
- Discussions about data acquisition, quality control, storage, retrieval and analyses
- Understanding concepts for operational time and trouble analyses, learning curve, technical limit and benchmarking
- Real business cases to illustrate the concepts explored in the course.

Pricing & Registration

SPE Members: USD 375 (Local), USD 750 (International) **SPE Non-members:** USD 450 (Local), USD 900 (International)

REGISTER FOR THIS COURSE

Learning Level

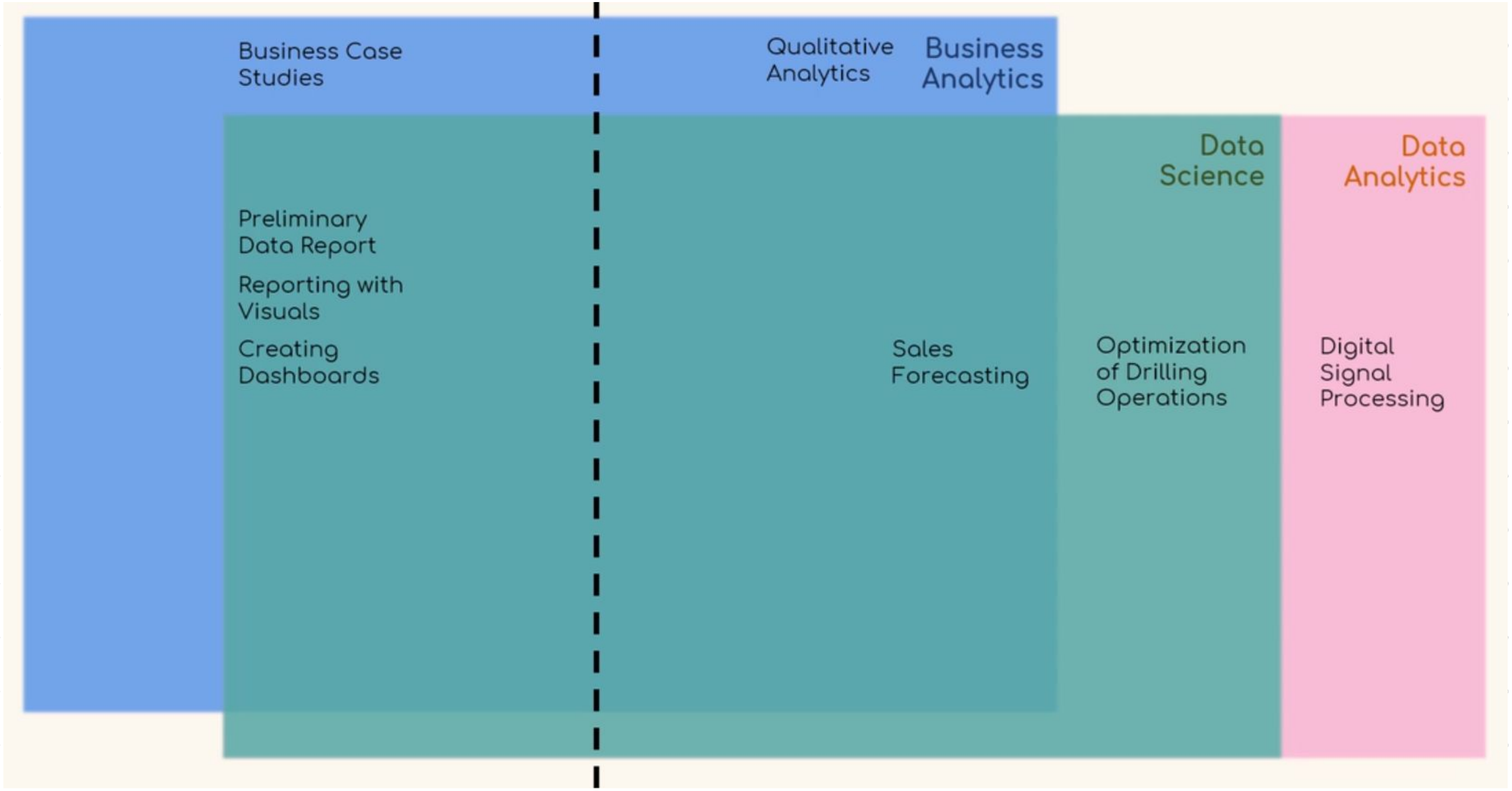
Introductory

Course Length

1 Day (Daily Activities Agenda)

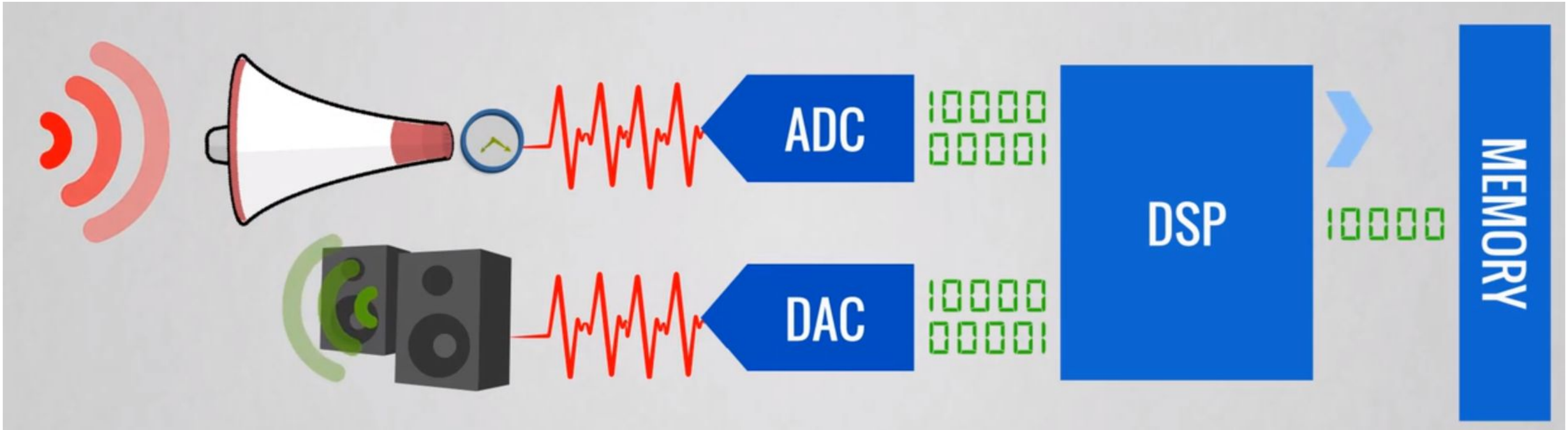
https://www.genesispetroleum.com.au/data-analytics-for-drilling-optimisation-cairo/

Digital Signal Processing (DSP) - 1



Digital Signal Processing (DSP) - 2

- Apply **Data Analytics** to **Digital Signal** to produce a higher quality signal
  - Analog-to-Digital Conversion (ADC)
  - Digital-to-Analog Conversion (DAC)

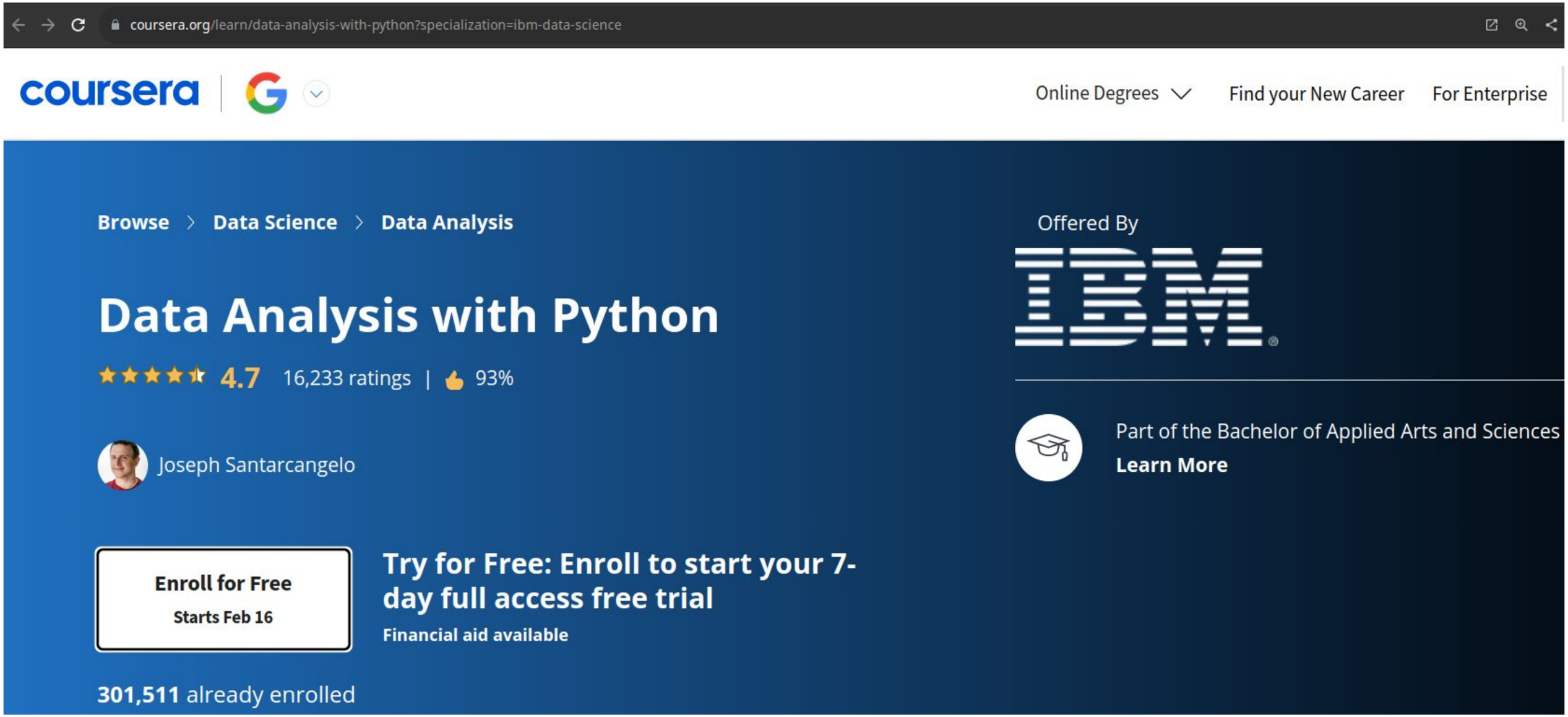


Coursera Data Analysis Course

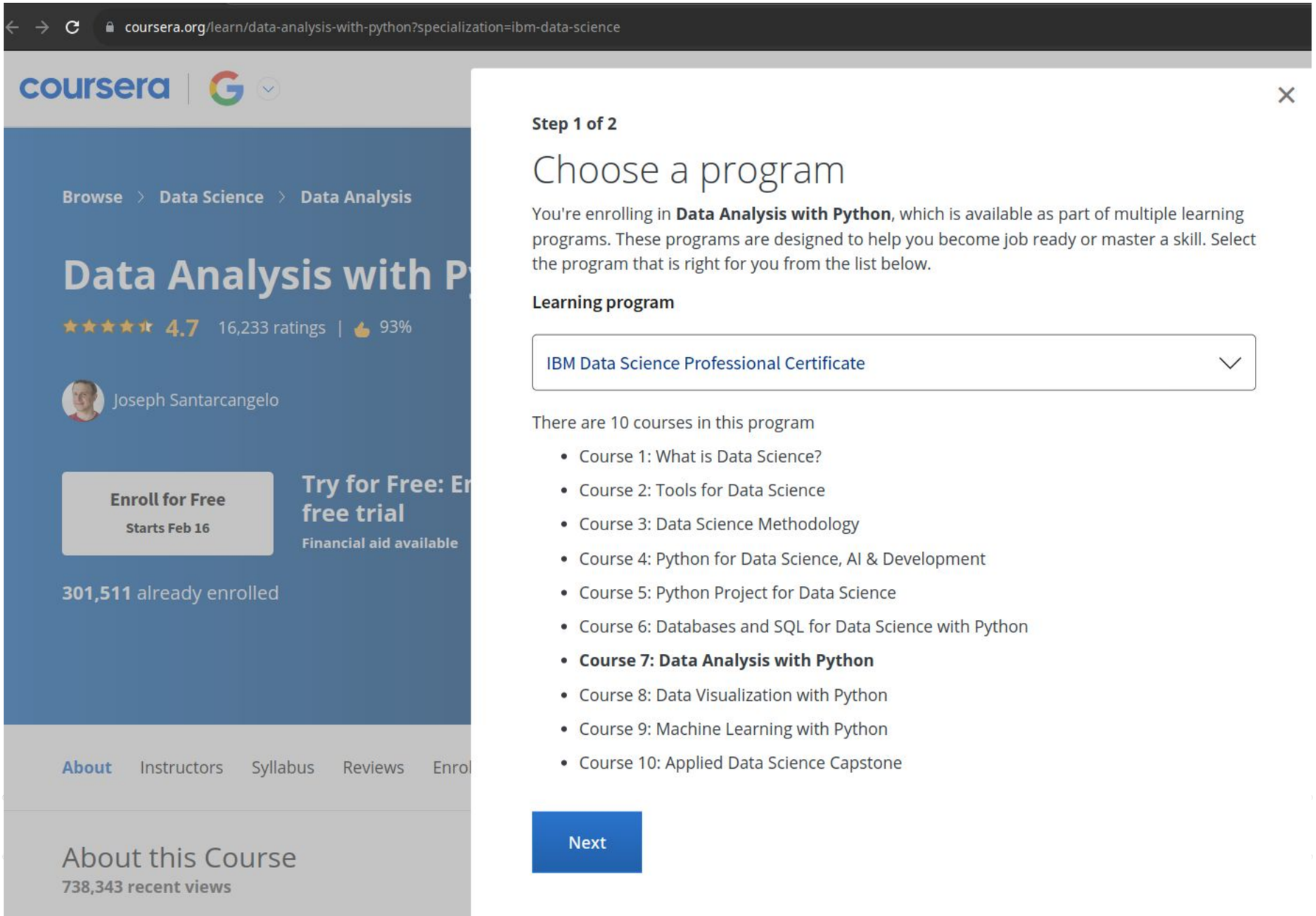
Coursera Data Analysis Course Link

https://www.coursera.org/learn/data-analysis-with-python?specialization=ibm-data-science

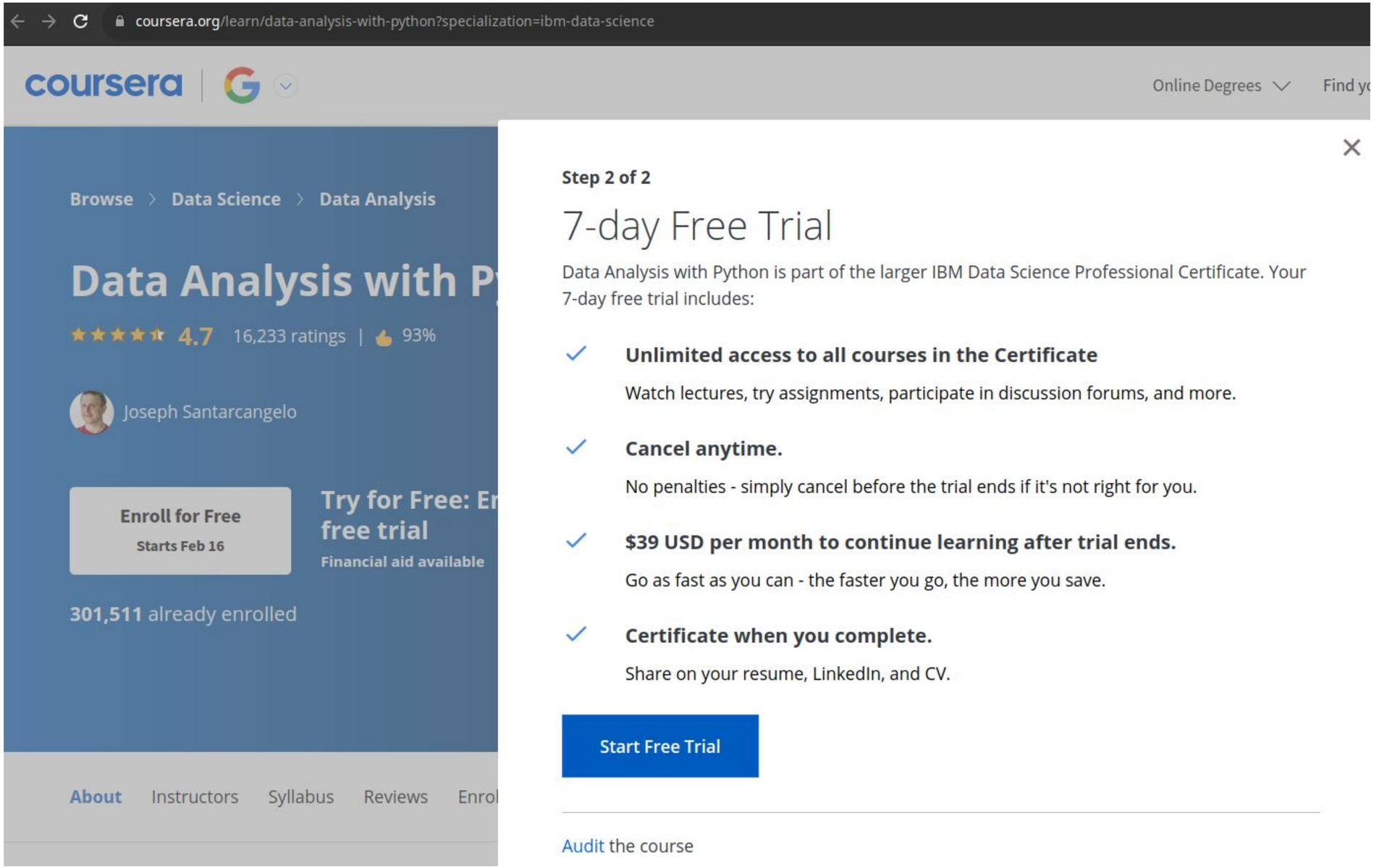
Enroll in Coursera Courses - 1



Enroll in Coursera Courses - 2



Enroll in Coursera Courses - 3



Python Packages

Why Data Analysis?

- Data is everywhere.
- Data analysis/data science helps us answer questions from data.
- **Data analysis** plays an important role in:
  - Discovering useful information
  - Answering questions
  - Predicting future or the unknown

Data Analysis Example: Sell a Car - 1

Tom wants to sell his car

Tom

How much money should he sell his car for?

The price he sets should not be too high, but not too low either.

Data Analysis Example: Sell a Car - 2

How can we help Tom determine the best price for his car?

- Is there data on the prices of other cars and their characteristics?
- What features of cars affect their prices?
  - Color? Brand? Horsepower? Something else?
- Asking the right questions in terms of data

Data Analysis Example: Used Cars Dataset - 1

```
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495
3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,16500
1,?,alfa-romero,gas,std,two,hatchback,rwd,front,94.50,171.20,65.50,52.40,2823,ohcv,six,152,mpfi,2.68,3.47,9.00,154,5000,19,26,16500
2,164,audi,gas,std,four,sedan,fwd,front,99.80,176.60,66.20,54.30,2337,ohc,four,109,mpfi,3.19,3.40,10.00,102,5500,24,30,13950
2,164,audi,gas,std,four,sedan,4wd,front,99.40,176.60,66.40,54.30,2824,ohc,five,136,mpfi,3.19,3.40,8.00,115,5500,18,22,17450
2,?,audi,gas,std,two,sedan,fwd,front,99.80,177.30,66.30,53.10,2507,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,15250
1,158,audi,gas,std,four,sedan,fwd,front,105.80,192.70,71.40,55.70,2844,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,17710
1,?,audi,gas,std,four,wagon,fwd,front,105.80,192.70,71.40,55.70,2954,ohc,five,136,mpfi,3.19,3.40,8.50,110,5500,19,25,18920
1,158,audi,gas,turbo,four,sedan,fwd,front,105.80,192.70,71.40,55.90,3086,ohc,five,131,mpfi,3.13,3.40,8.30,140,5500,17,20,23875
0,?,audi,gas,turbo,two,hatchback,4wd,front,99.50,178.20,67.90,52.00,3053,ohc,five,131,mpfi,3.13,3.40,7.00,160,5500,16,22,?
2,192,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16430
0,192,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2395,ohc,four,108,mpfi,3.50,2.80,8.80,101,5800,23,29,16925
0,188,bmw,gas,std,two,sedan,rwd,front,101.20,176.80,64.80,54.30,2710,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,20970
0,188,bmw,gas,std,four,sedan,rwd,front,101.20,176.80,64.80,54.30,2765,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,21,28,21105
1,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3055,ohc,six,164,mpfi,3.31,3.19,9.00,121,4250,20,25,24565
0,?,bmw,gas,std,four,sedan,rwd,front,103.50,189.00,66.90,55.70,3230,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,30760
0,?,bmw,gas,std,two,sedan,rwd,front,103.50,193.80,67.90,53.70,3380,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,16,22,41315
0,?,bmw,gas,std,four,sedan,rwd,front,110.00,197.00,70.90,56.30,3505,ohc,six,209,mpfi,3.62,3.39,8.00,182,5400,15,20,36880
2,121,chevrolet,gas,std,two,hatchback,fwd,front,88.40,141.10,60.30,53.20,1488,l,three,61,2bbl,2.91,3.03,9.50,48,5100,47,53,5151
```

Data Analysis Example: Used Cars Dataset - 2

No.	Attribute name	attribute range	No.	Attribute name	attribute range
1	symboling	-3, -2, -1, 0, 1, 2, 3.	14	curb-weight	continuous from 1488 to 4066.
2	normalized-losses	continuous from 65 to 256.	15	engine-type	dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
3	make	audi, bmw, etc.	16	num-of-cylinders	eight, five, four, six, three, twelve, two.
4	fuel-type	diesel, gas.	17	engine-size	continuous from 61 to 326.
5	aspiration	std, turbo.	18	fuel-system	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
6	num-of-doors	four, two.	19	bore	continuous from 2.54 to 3.94.
7	body-style	hardtop, wagon, etc.	20	stroke	continuous from 2.07 to 4.17.
8	drive-wheels	4wd, fwd, rwd.	21	compression-ratio	continuous from 7 to 23.
9	engine-location	front, rear.	22	horsepower	continuous from 48 to 288.
10	wheel-base	continuous from 86.6 120.9.	23	peak-rpm	continuous from 4150 to 6600.
11	length	continuous from 141.1 to 208.1.	24	city-mpg	continuous from 13 to 49.
12	width	continuous from 60.3 to 72.3.	25	highway-mpg	continuous from 16 to 54.
13	height	continuous from 47.8 to 59.8.	26	price	continuous from 5118 to 45400.

Data Analysis Example: Used Cars Dataset - 3

https://archive.ics.uci.edu/ml/machine-learning-databases/autos/

Scientific Computing Libraries in Python

1. **Scientifics Computing** Libraries

Pandas
(Data structures & tools)

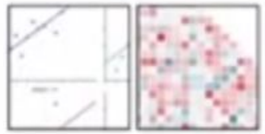
NumPy
(Arrays & matrices)

SciPy
(Integrals, solving differential equations, optimization)

Visualization Libraries in Python

2. **Visualization** Libraries

Matplotlib
(plots & graphs, most popular)

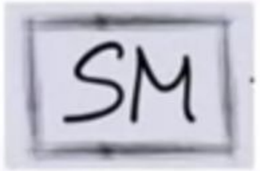


Seaborn
(plots : heat maps, time series, violin plots)

Algorithmic Libraries in Python

3. **Algorithmic** libraries

Scikit-learn
(Machine Learning : regression, classification,…)

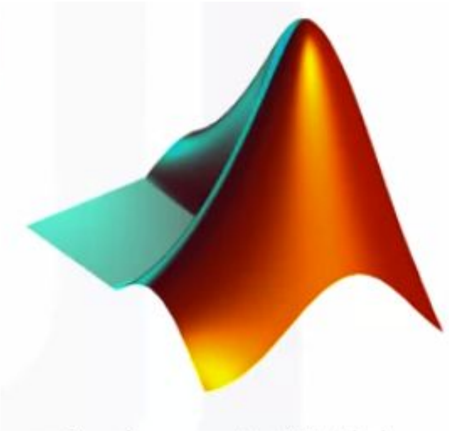


Statsmodels
(Explore data, estimate statistical models, and perform statistical tests.)

John Hunter (Matplotlib Creator)

- Neurobiologist
- Part of a team analyzing **Electrocorticography Signals (ECoG)**
  - **Electrocorticography** is the process of recording electrical activity in the brain

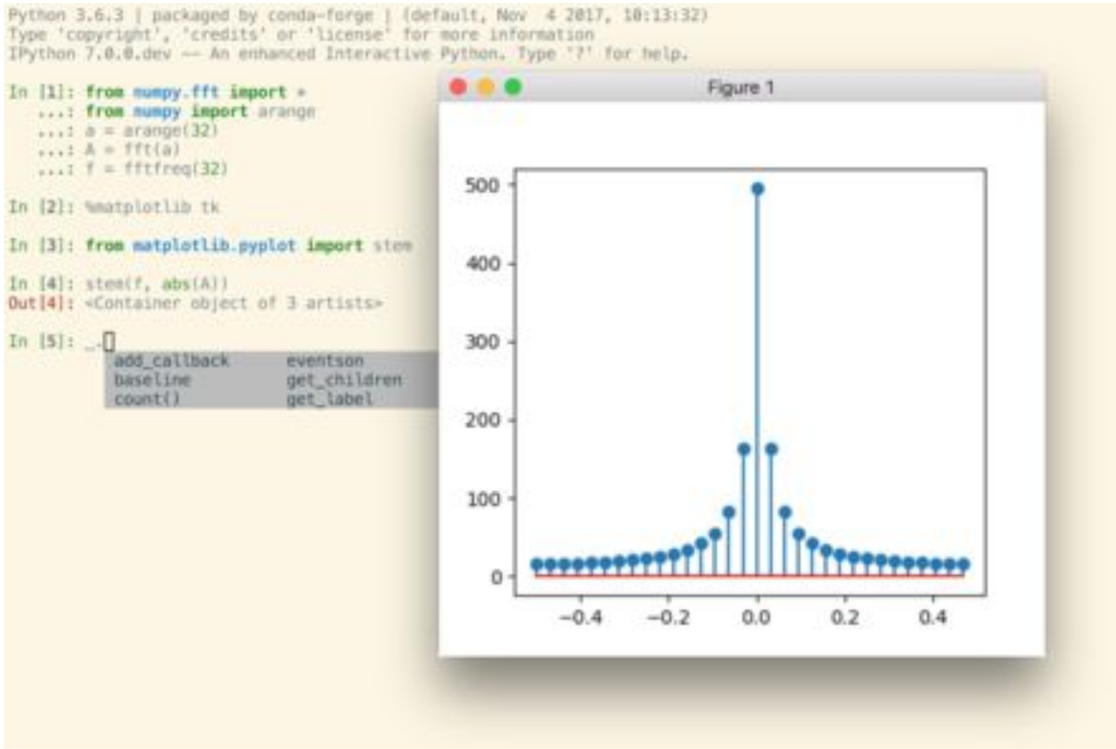
- The team
  - used a proprietary software (**MATLAB** based version) for analysis
  - had only one license and were taking turns in using it



- **John** replace the proprietary software with **Matplotlib**

Jupyter Notebook

- open source web app
- allows to create & share documents
  - that contain code and text
- spun off from **iPython** in **2014**



- **Jupyter name** is a reference to three programming languages:
  - **Ju**lia
  - **Pyt**hon
  - **R**
- **Jupyter logo**
  - homage to **Galileo**'s discovery of the **moons of Jupiter**
  - documented in **notebooks** attributed to **Galileo**

Questions

Links

https://github.com/fcai-b/da

References

1. https://learn.365datascience.com/courses/intro-to-data-and-data-science
    - 365 Data Science - Introduction to Data and Data Science
2. https://www.coursera.org/learn/data-analysis-with-python
    - IBM Coursera Course - Data Analysis with Python