

Introduction to Machine Learning

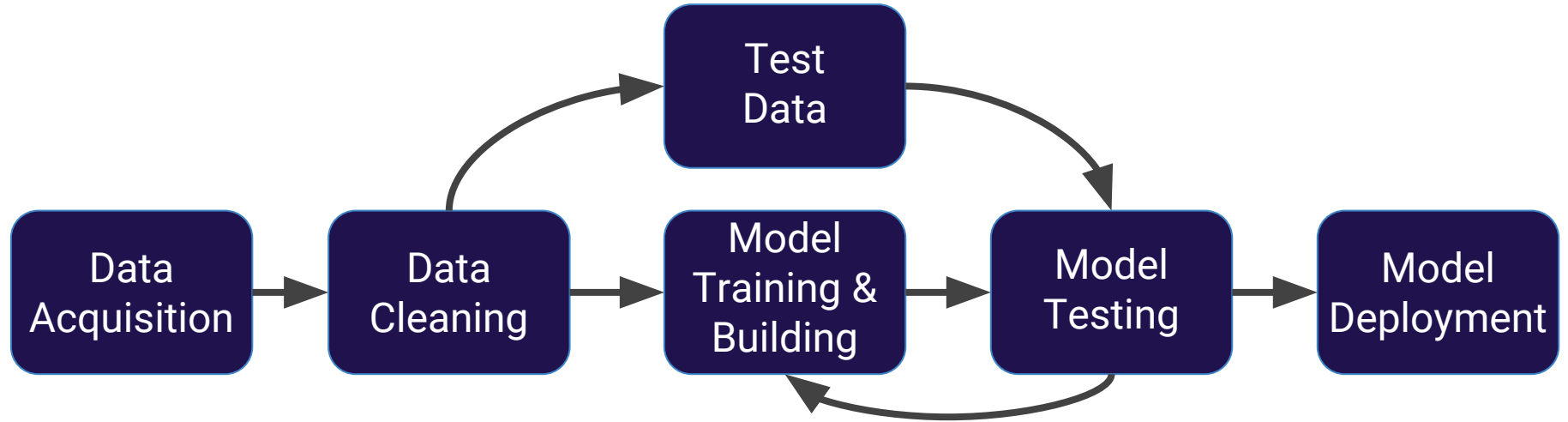
What is Machine Learning?

- Machine learning is a method of data analysis that automates analytical model building.
- Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

What is it used for?

- Fraud detection.
- Web search results.
- Real-time ads on web pages
- Credit scoring and next-best offers.
- Prediction of equipment failures.
- New pricing models.
- Network intrusion detection.
- Recommendation Engines
- Customer Segmentation
- Text Sentiment Analysis
- Predicting Customer Churn
- Pattern and image recognition.
- Email spam filtering.
- Financial Modeling

Machine Learning Process



Supervised Learning

- **Supervised learning** algorithms are trained using **labeled** examples, such as an input where the desired output is known.
- For example, a piece of equipment could have data points labeled either “F” (failed) or “R” (runs).

Supervised Learning

- The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors.
- It then modifies the model accordingly.

Supervised Learning

- Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.
- Supervised learning is commonly used in applications where historical data predicts likely future events.

Supervised Learning

- For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim.
- Or it can attempt to predict the price of a house based on different features for houses for which we have historical price data.

Unsupervised Learning

- **Unsupervised learning** is used against data that has no historical labels.
- The system is not told the "right answer." The algorithm must figure out what is being shown.
- The goal is to explore the data and find some structure within.

Unsupervised Learning

- Or it can find the main attributes that separate customer segments from each other.
- Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition.

Unsupervised Learning

- These algorithms are also used to segment text topics, recommend items and identify data outliers.

Reinforcement Learning

- **Reinforcement learning** is often used for robotics, gaming and navigation.
- With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards.

Reinforcement Learning

- This type of learning has three primary components: the agent (the learner or decision maker), the environment (everything the agent interacts with) and actions (what the agent can do).

Reinforcement Learning

- The objective is for the agent to choose actions that maximize the expected reward over a given amount of time.
- The agent will reach the goal much faster by following a good policy.

Reinforcement Learning

- So the goal in reinforcement learning is to learn the best policy.

Business Case
Studies

Qualitative
Analytics

Preliminary
Data Report

Reporting with
Visuals

Creating
Dashboards

Sales
Forecasting

Business

Business Case
Studies

Qualitative
Analytics

Preliminary
Data Report

Reporting with
Visuals

Creating
Dashboards

Sales
Forecasting

Business

Business Case
Studies

Business

Qualitative
Analytics

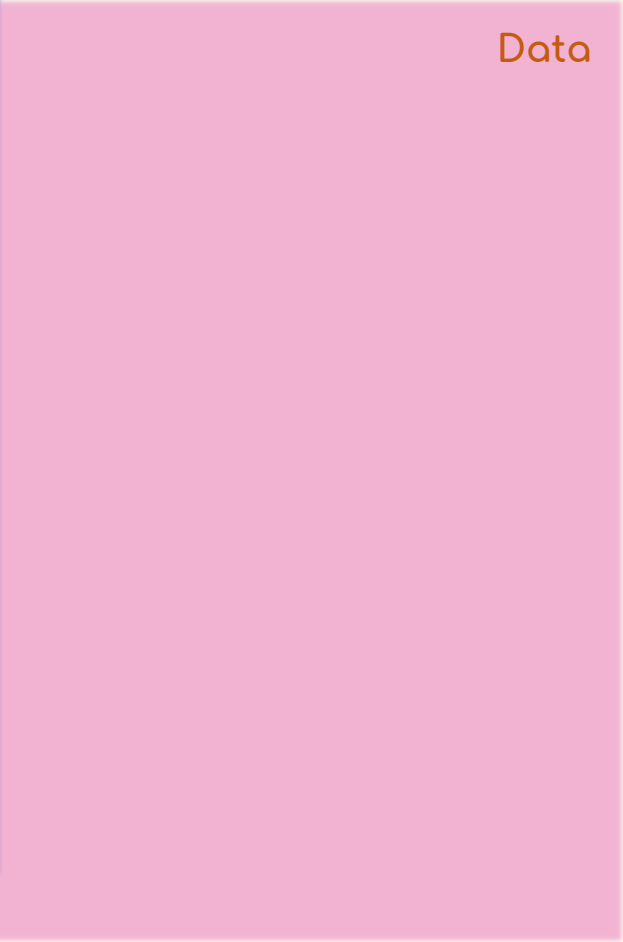
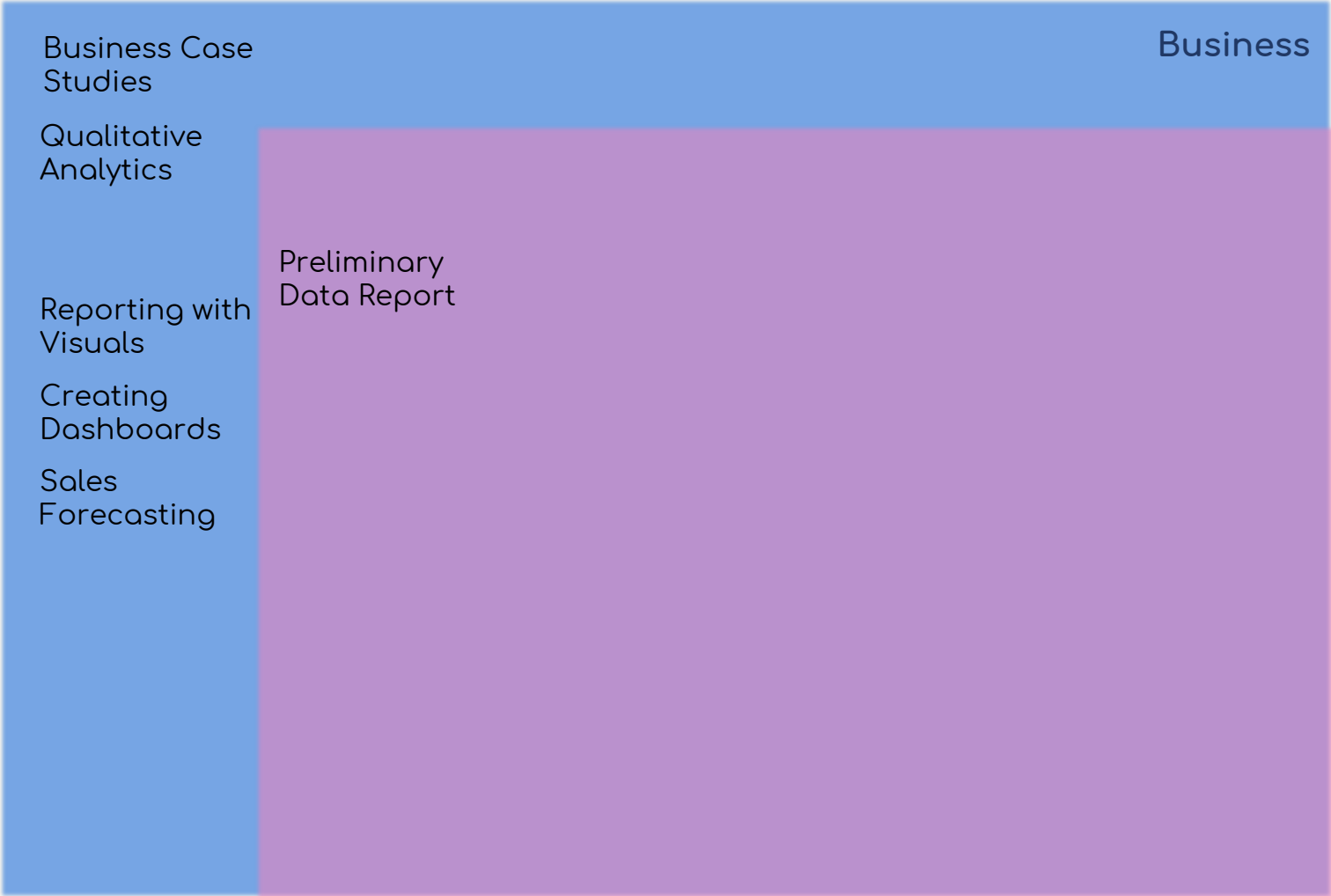
Preliminary
Data Report

Reporting with
Visuals

Creating
Dashboards

Sales
Forecasting

Data



Business Case
Studies

Qualitative
Analytics

Creating
Dashboards

Sales
Forecasting

Business

Preliminary
Data Report

Reporting with
Visuals

Data



Business Case
Studies

Business

Qualitative
Analytics

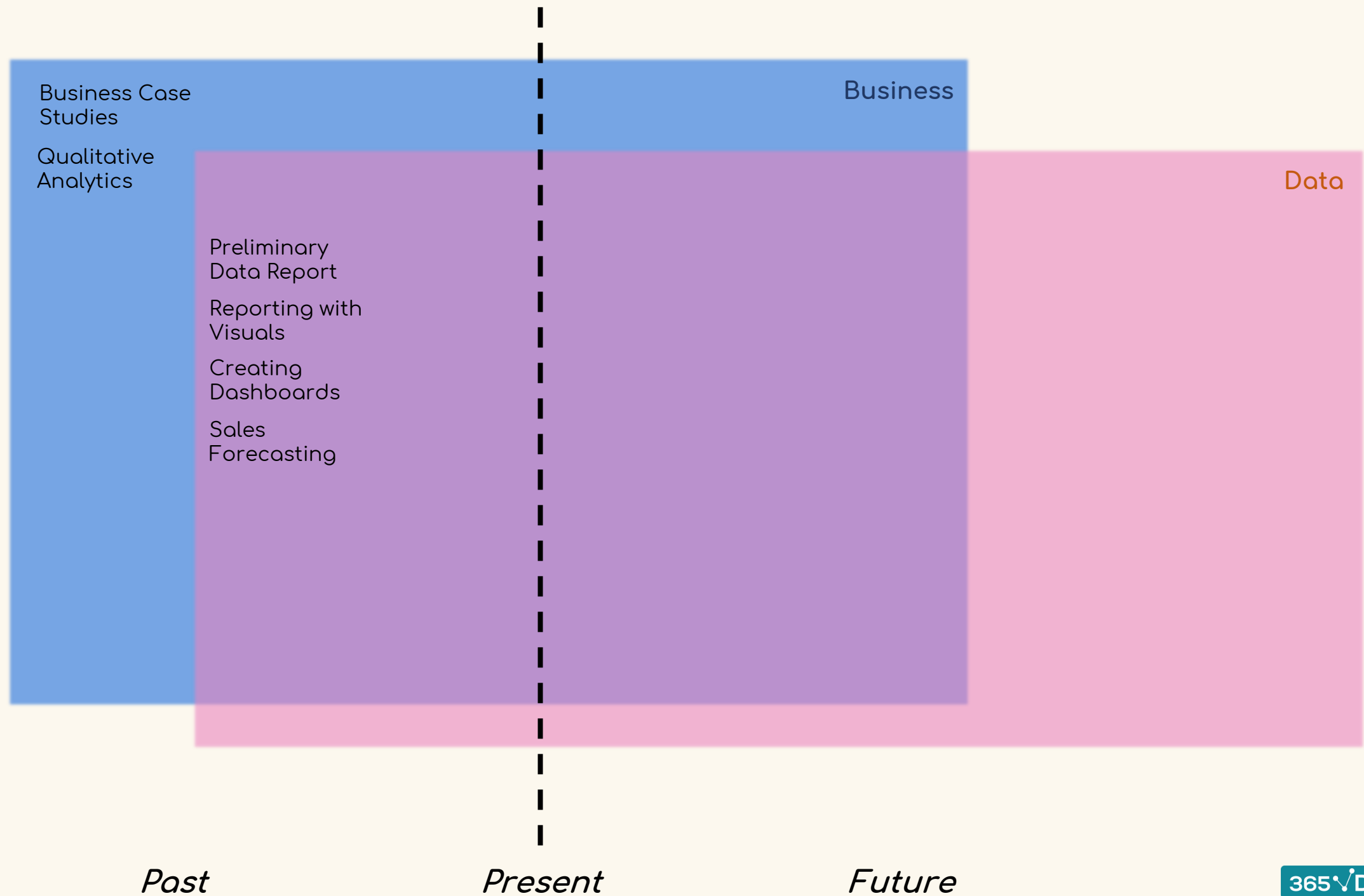
Data

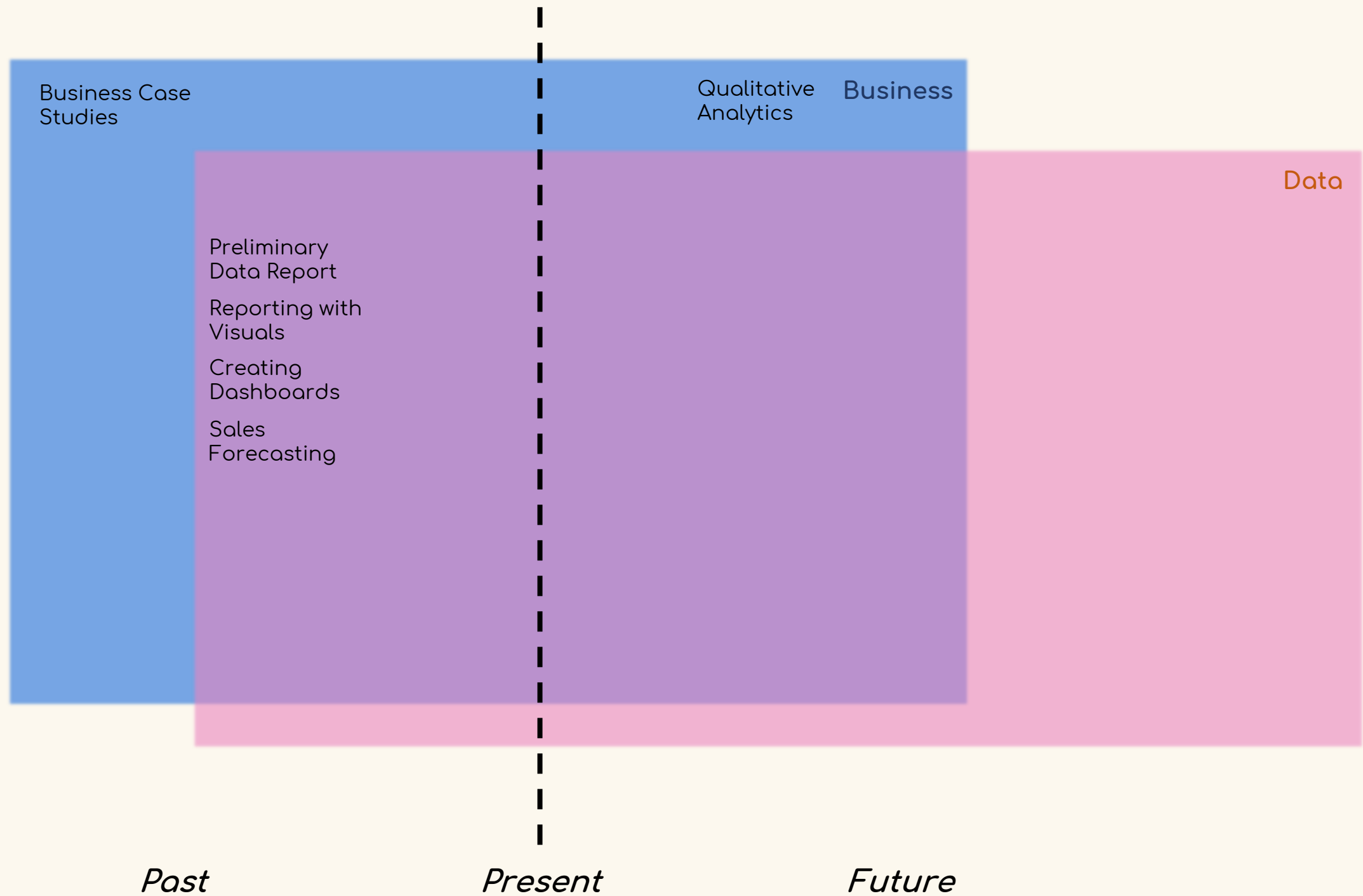
Preliminary
Data Report

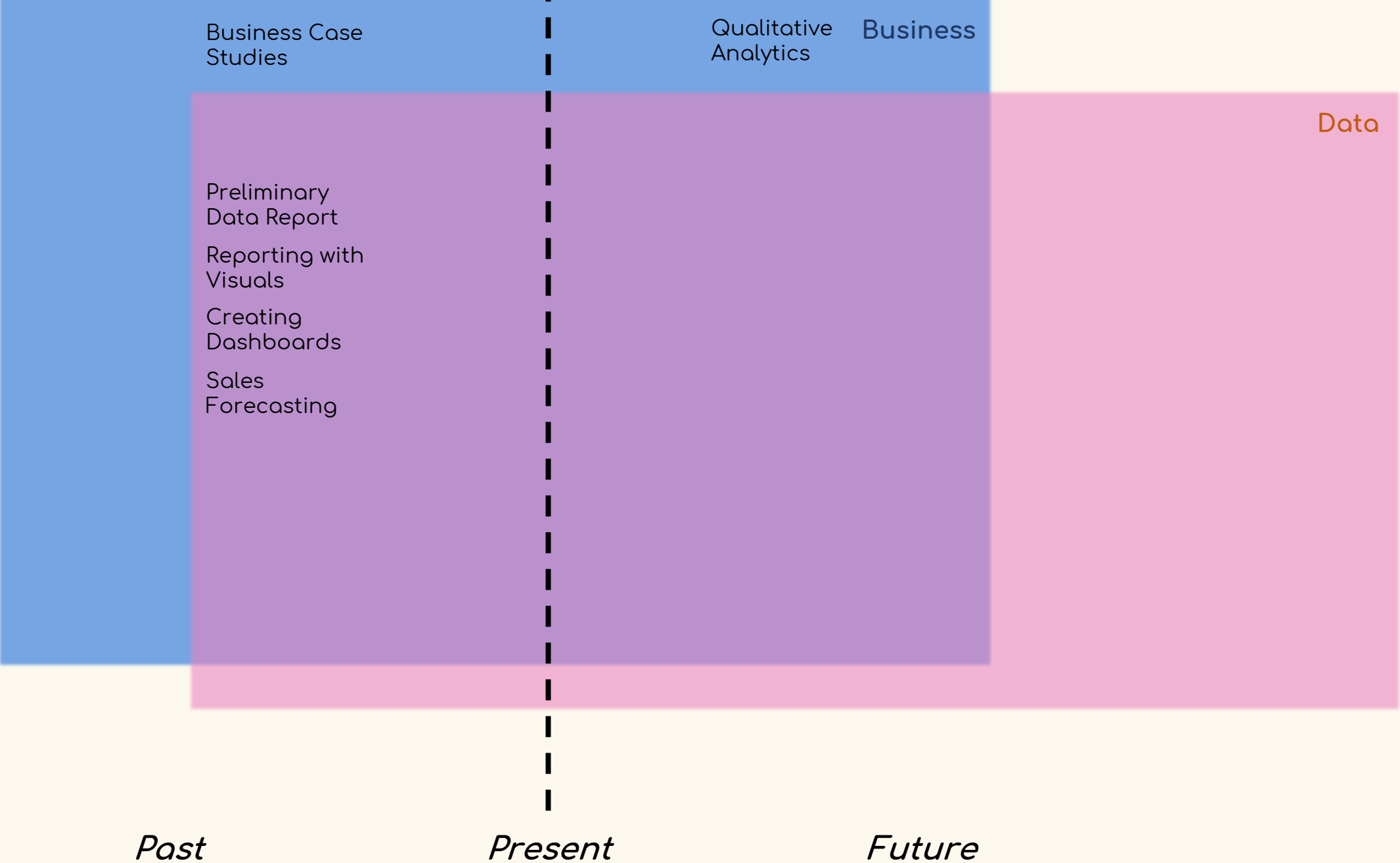
Reporting with
Visuals

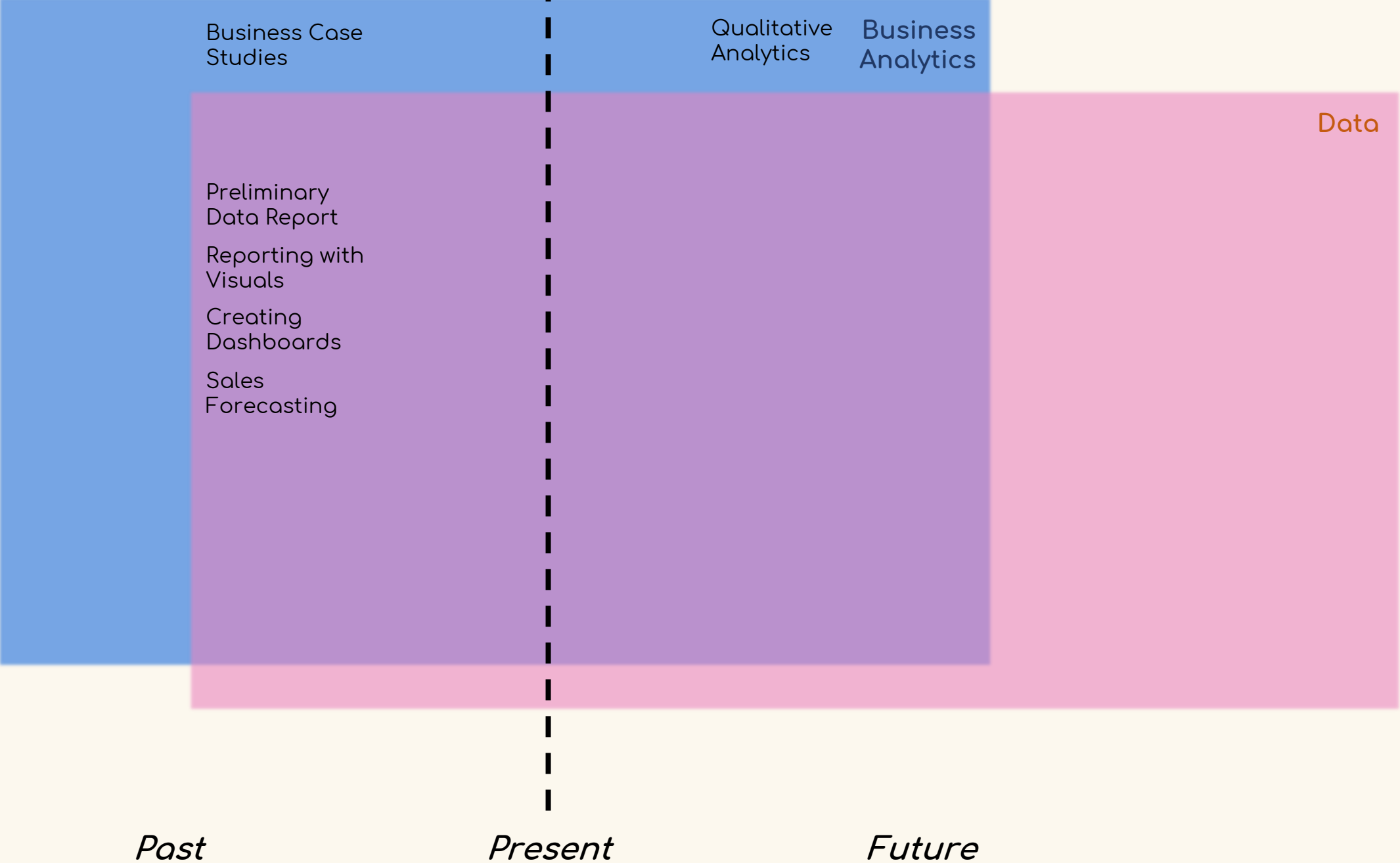
Creating
Dashboards

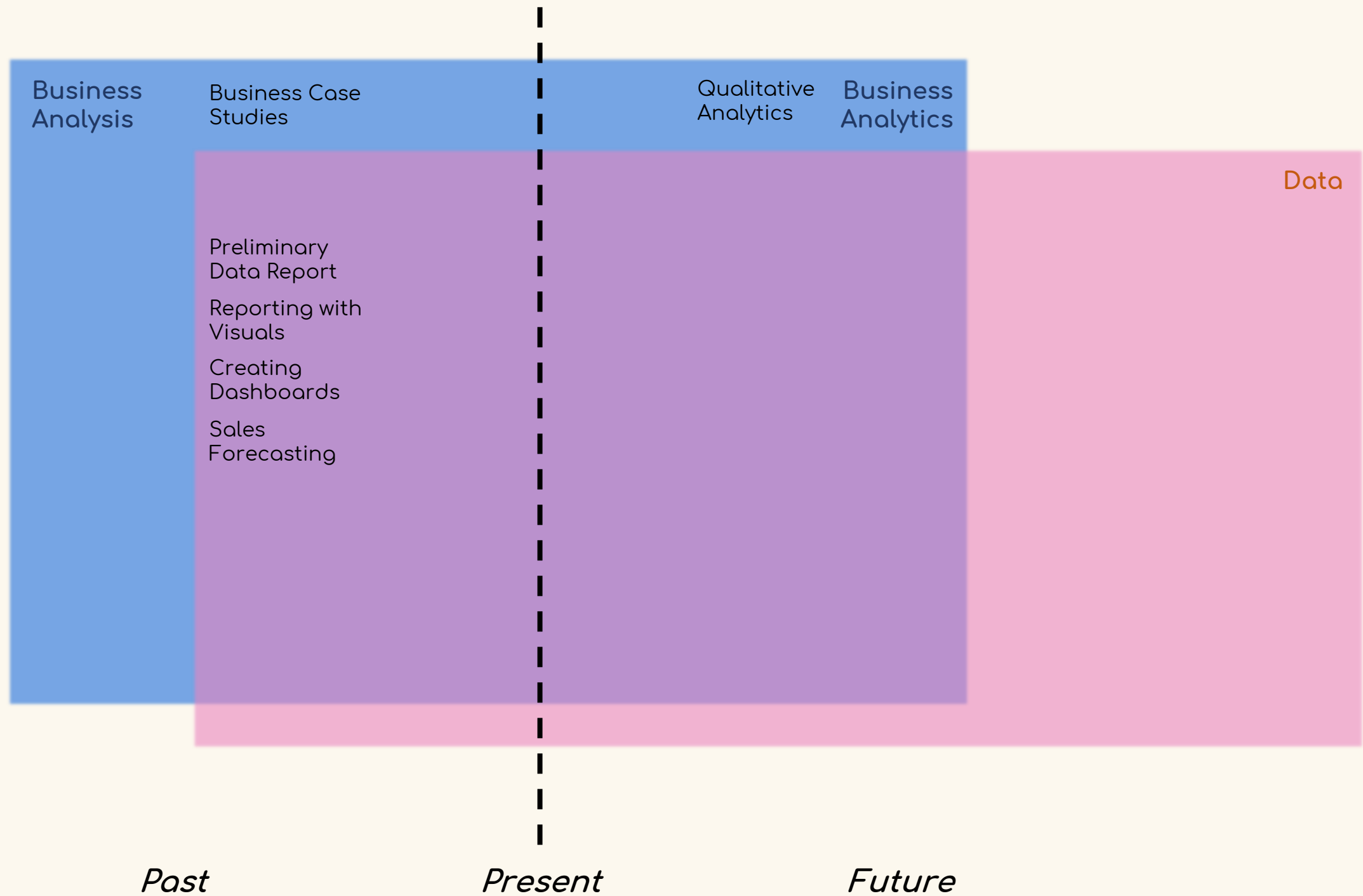
Sales
Forecasting

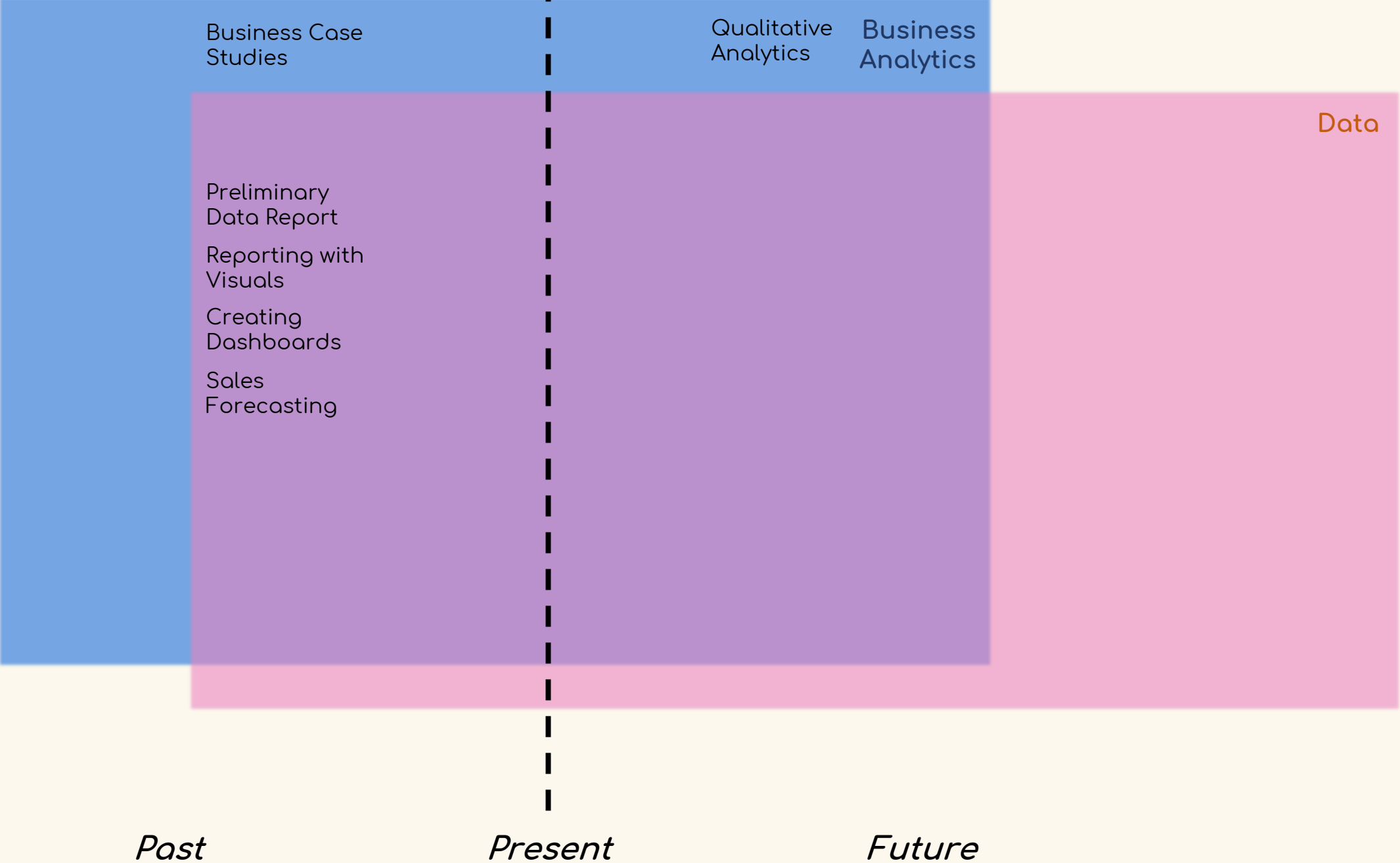


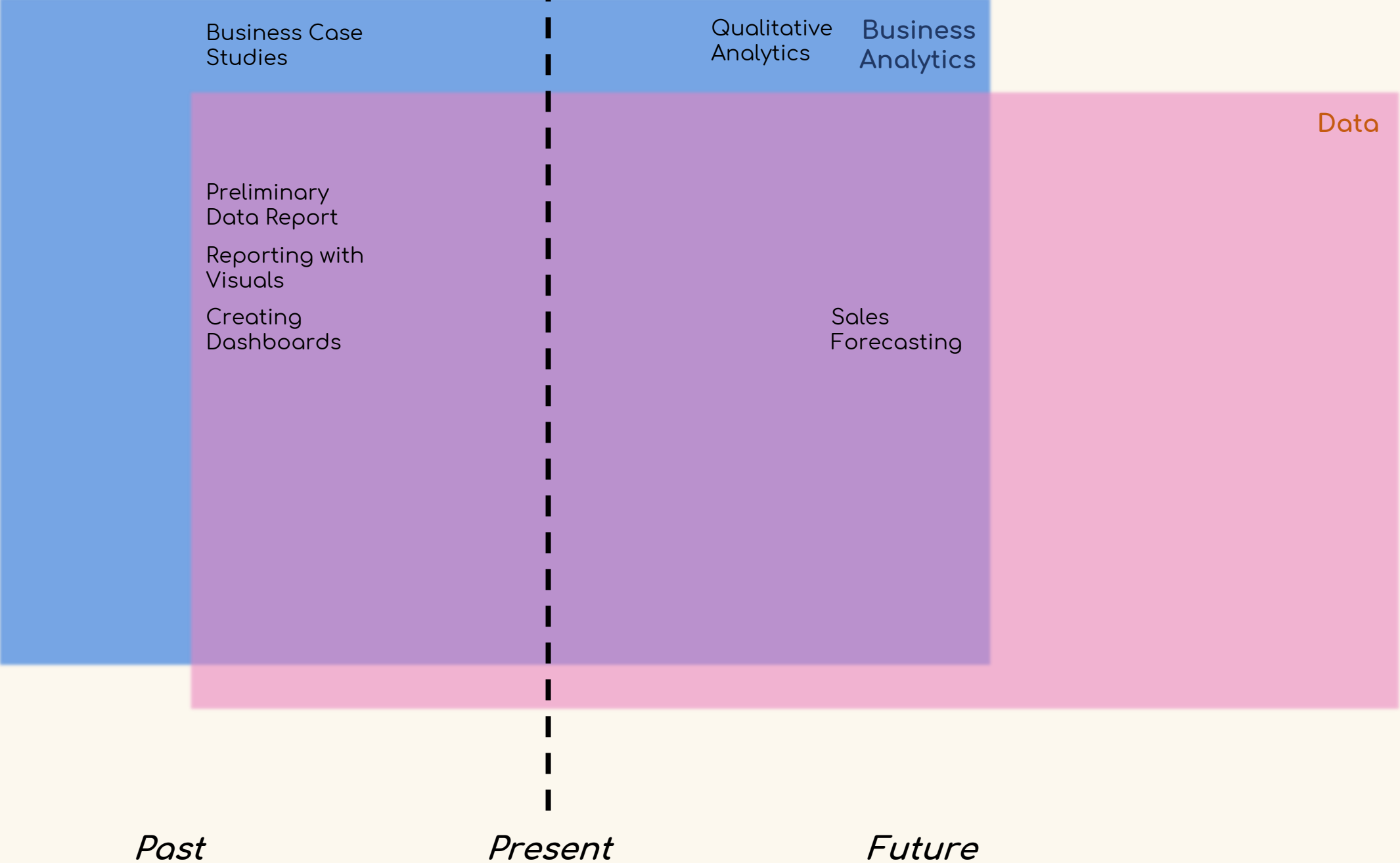


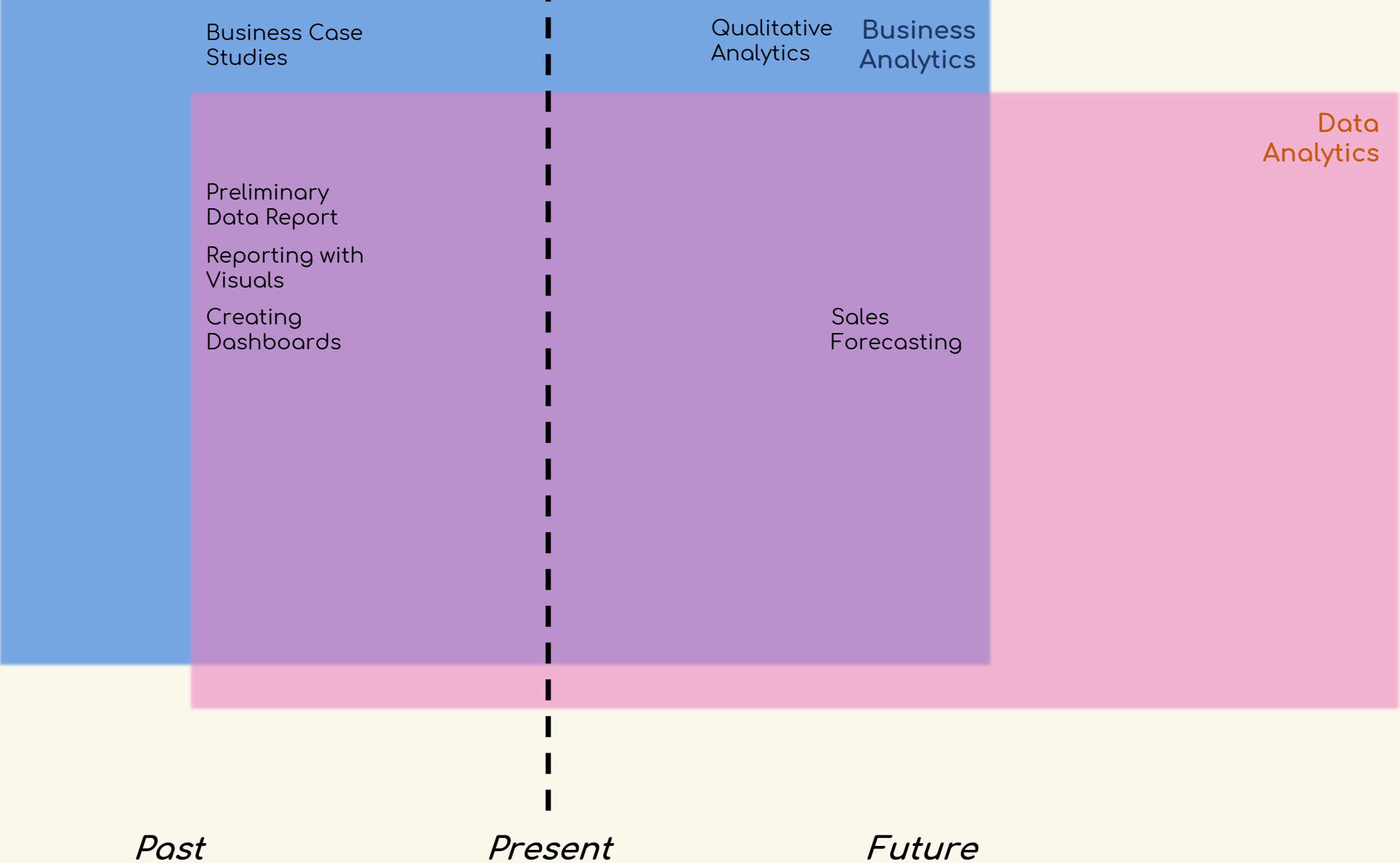


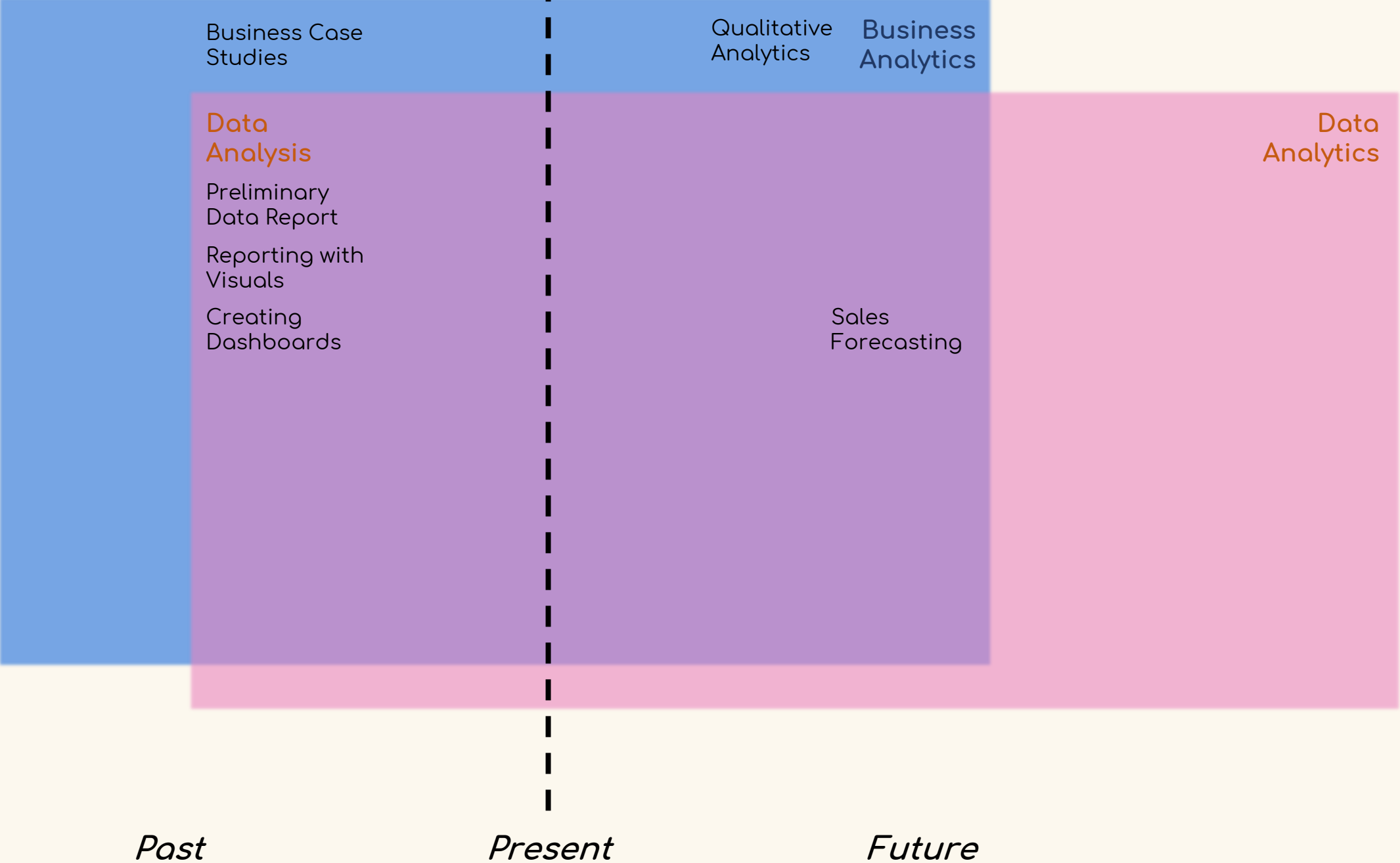


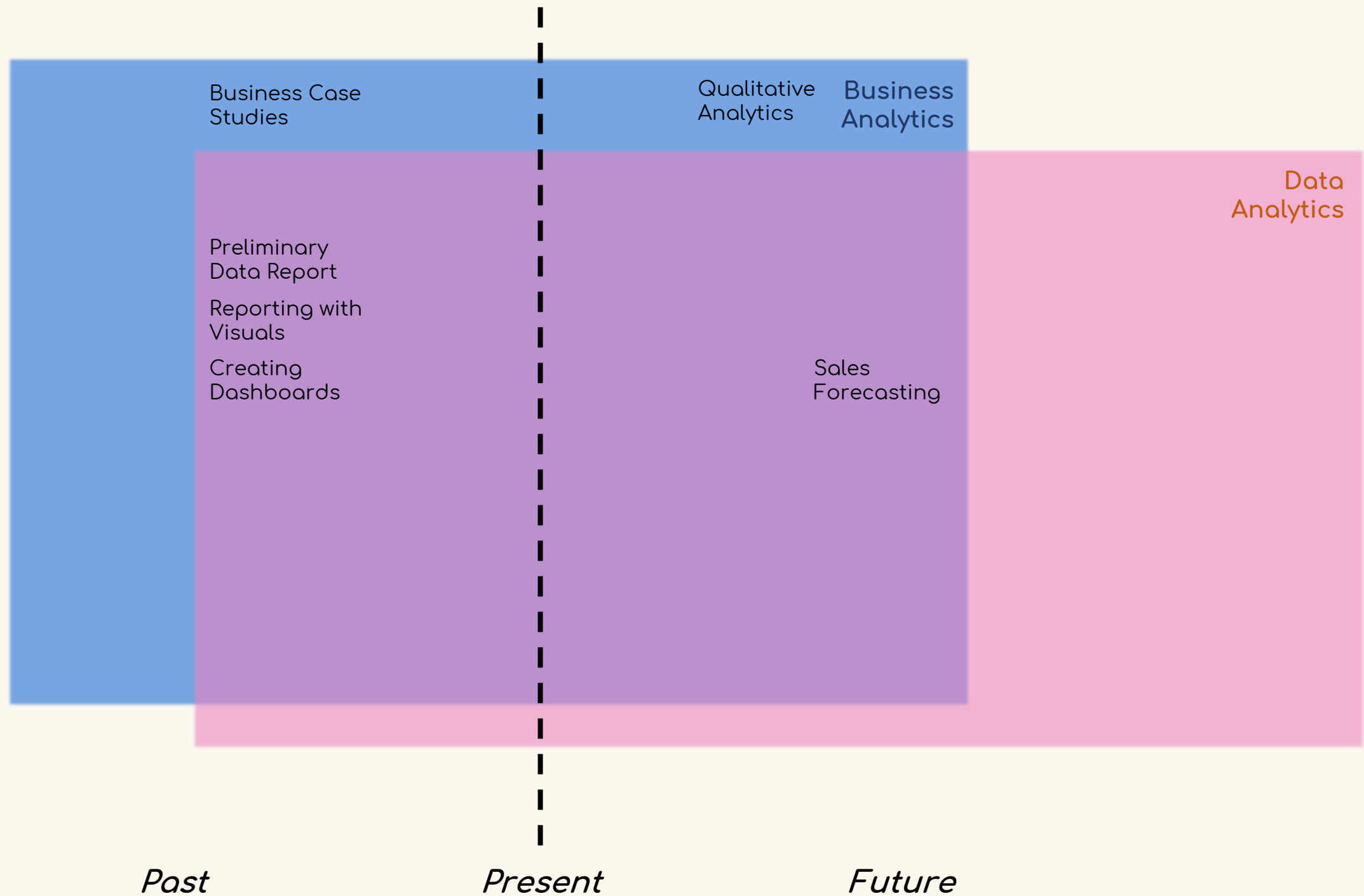


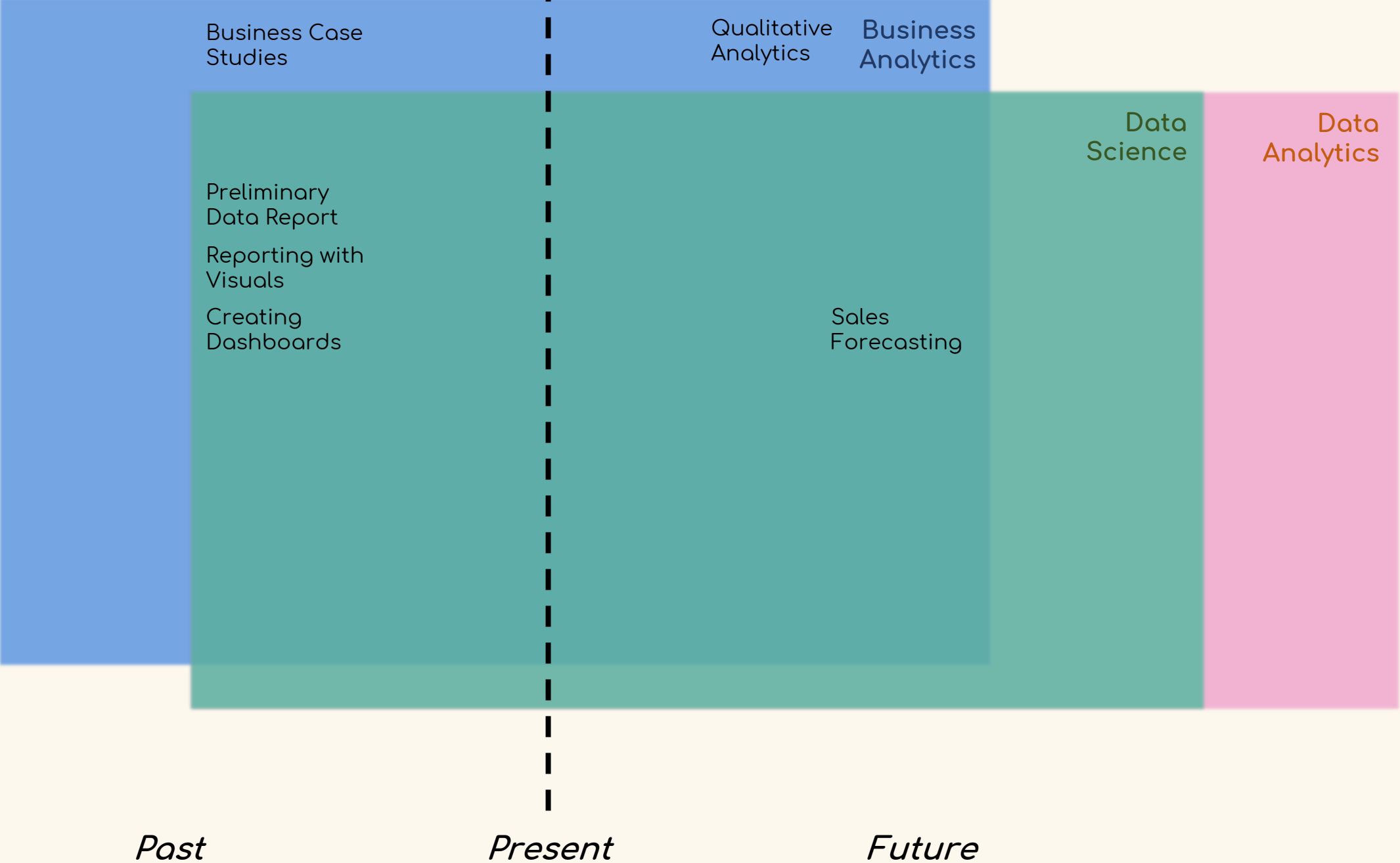


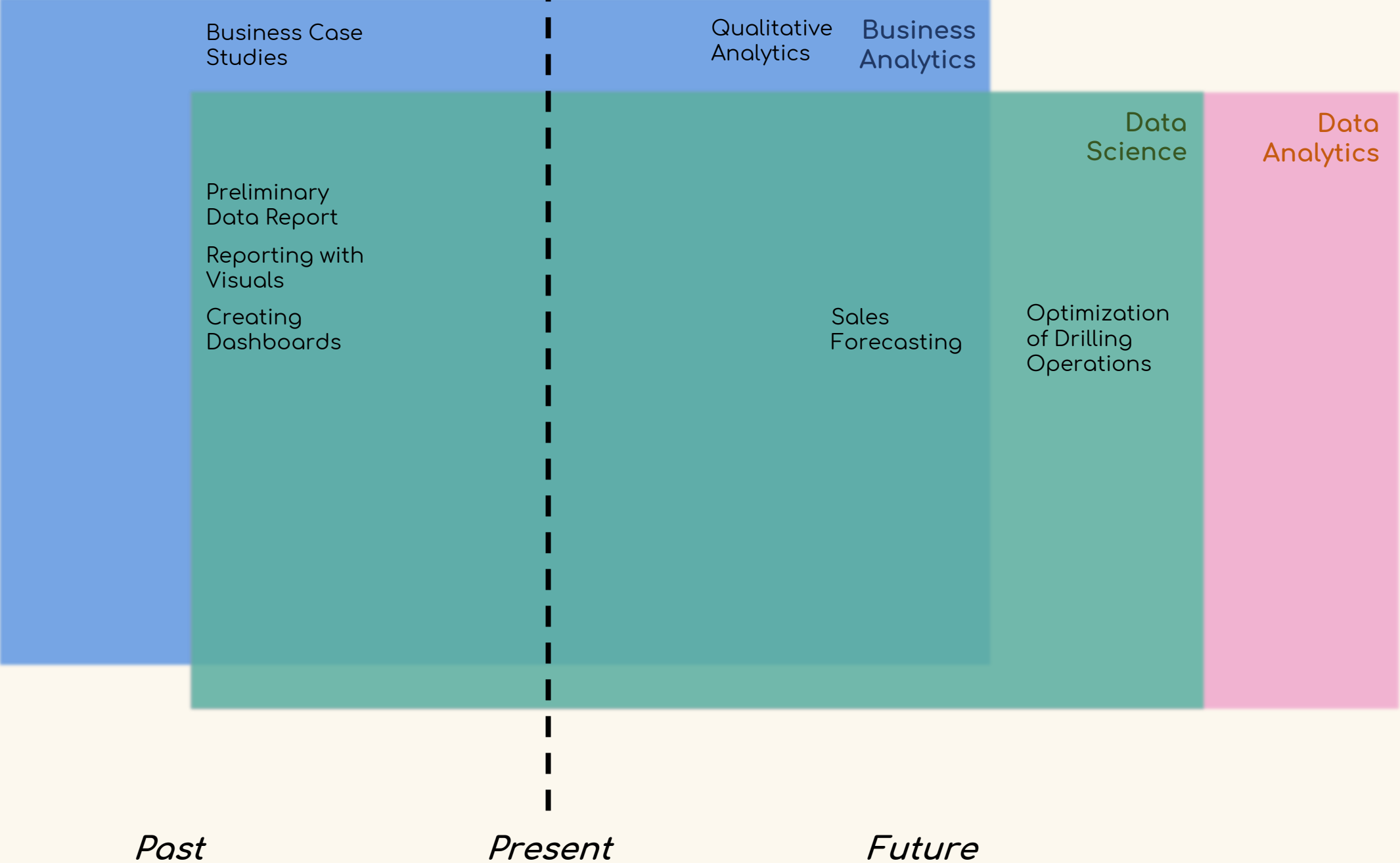


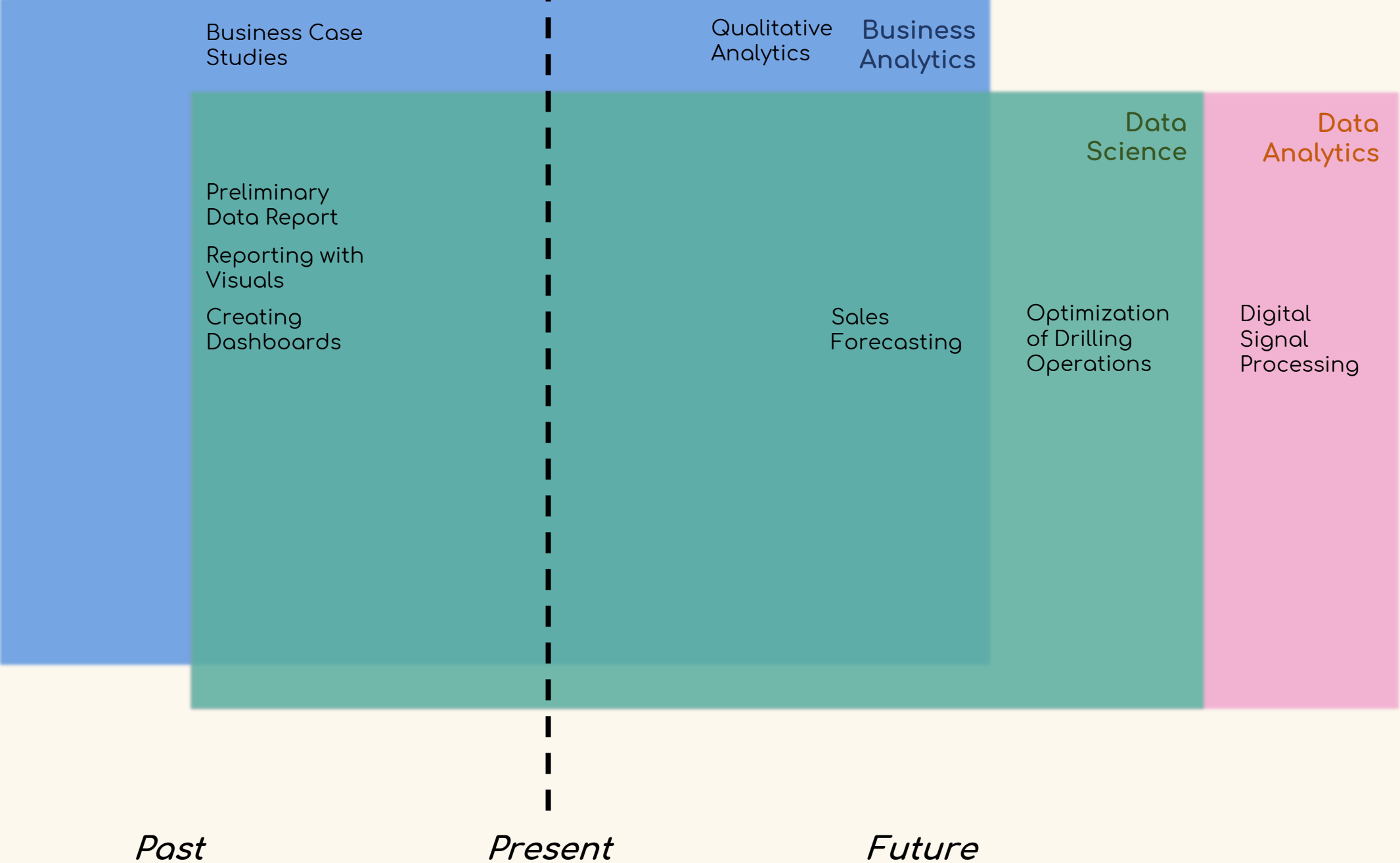


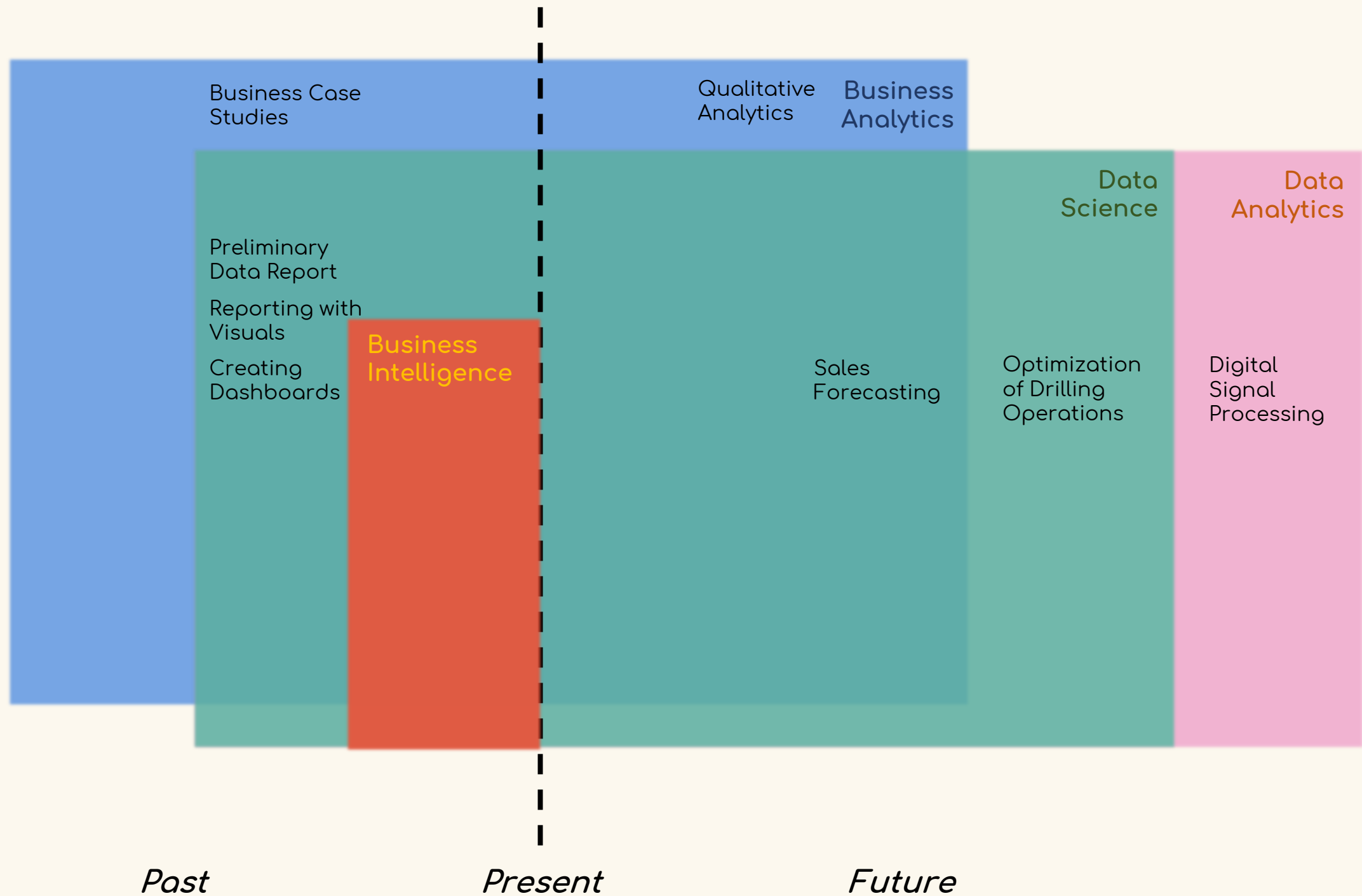


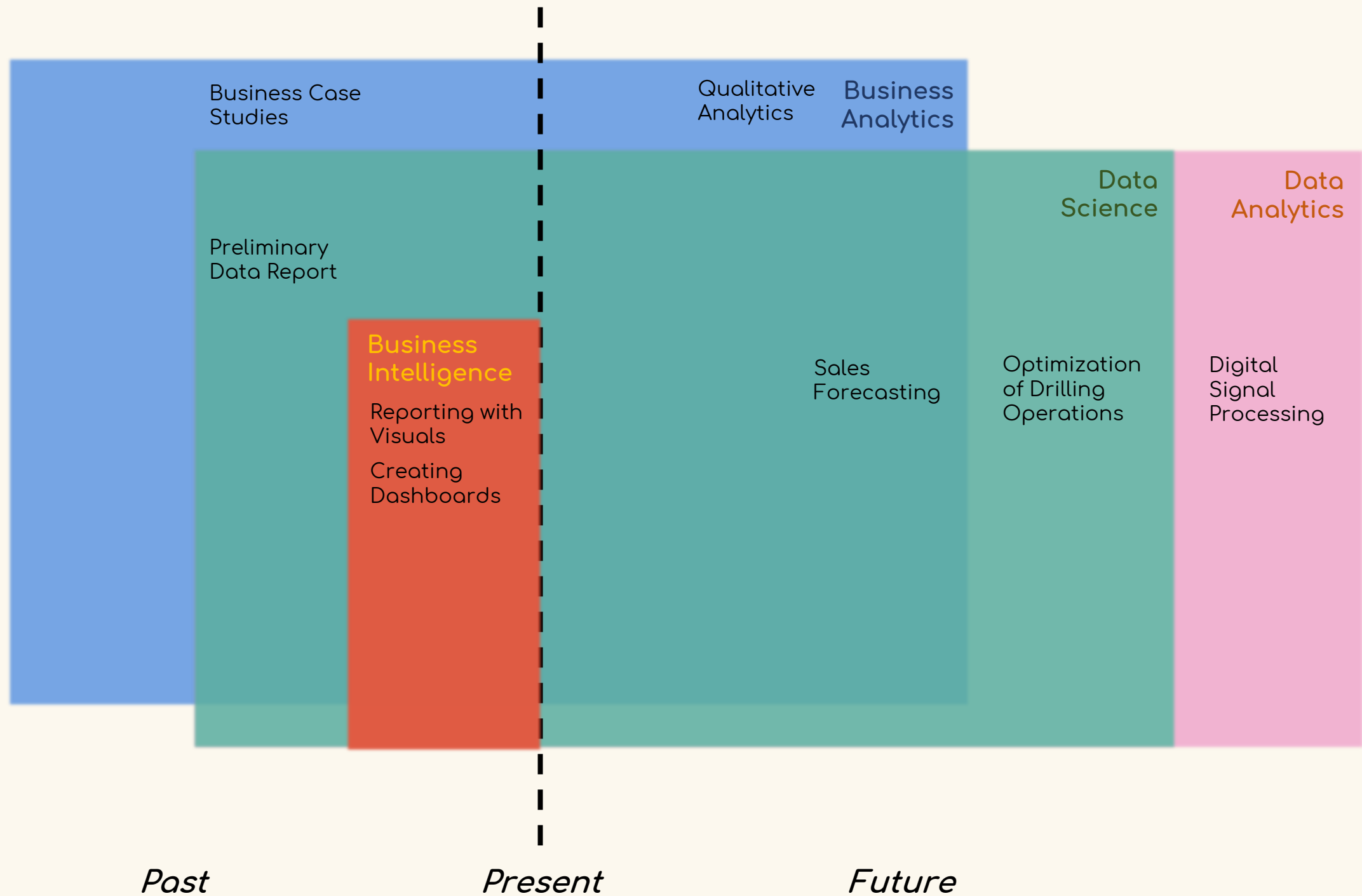


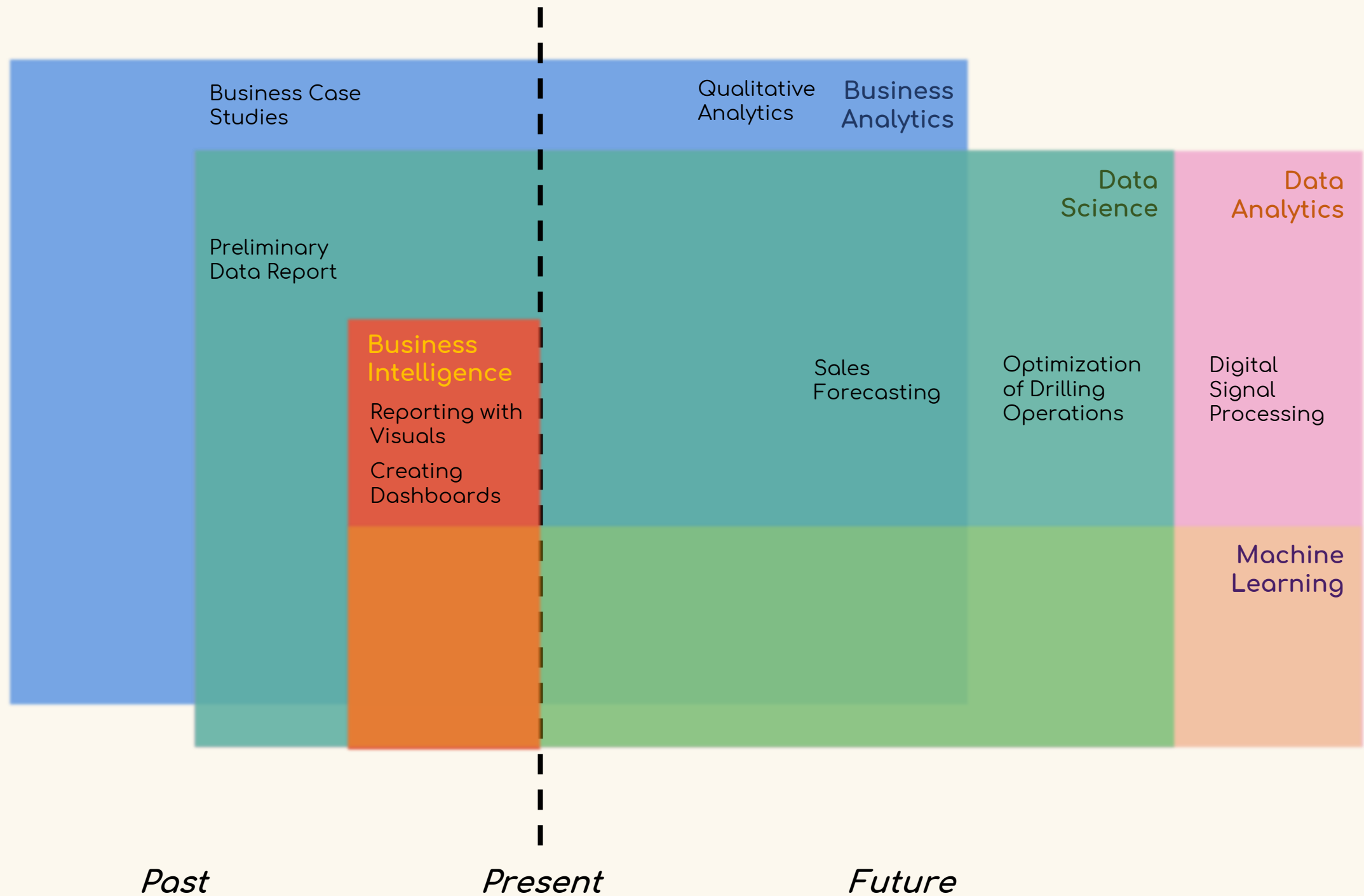


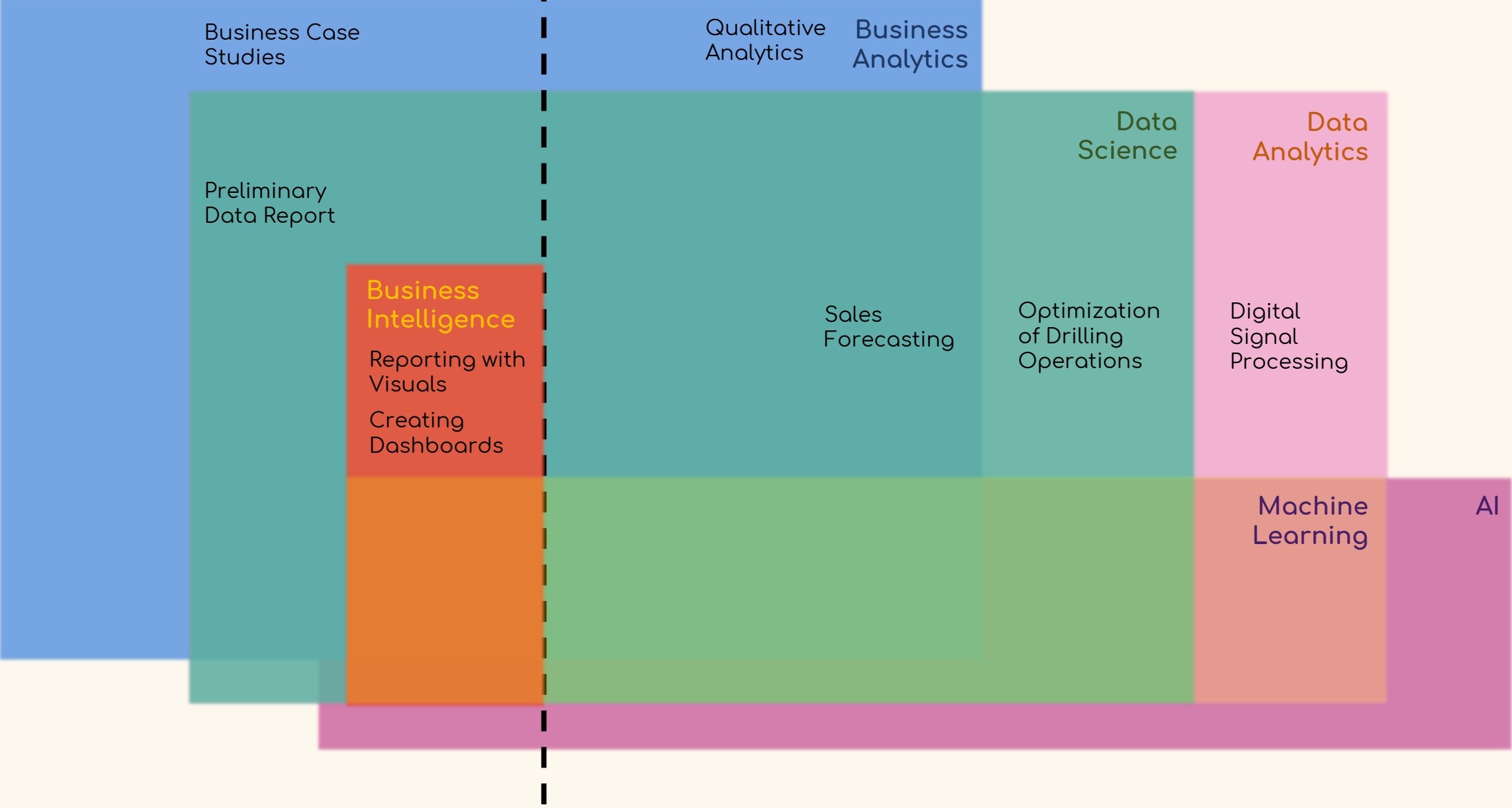












Business Case
Studies

Qualitative
Analytics

Business
Analytics

Preliminary
Data Report

Data
Science

Data
Analytics

Business
Intelligence

Reporting with
Visuals
Creating
Dashboards

Sales
Forecasting

Optimization
of Drilling
Operations

Digital
Signal
Processing

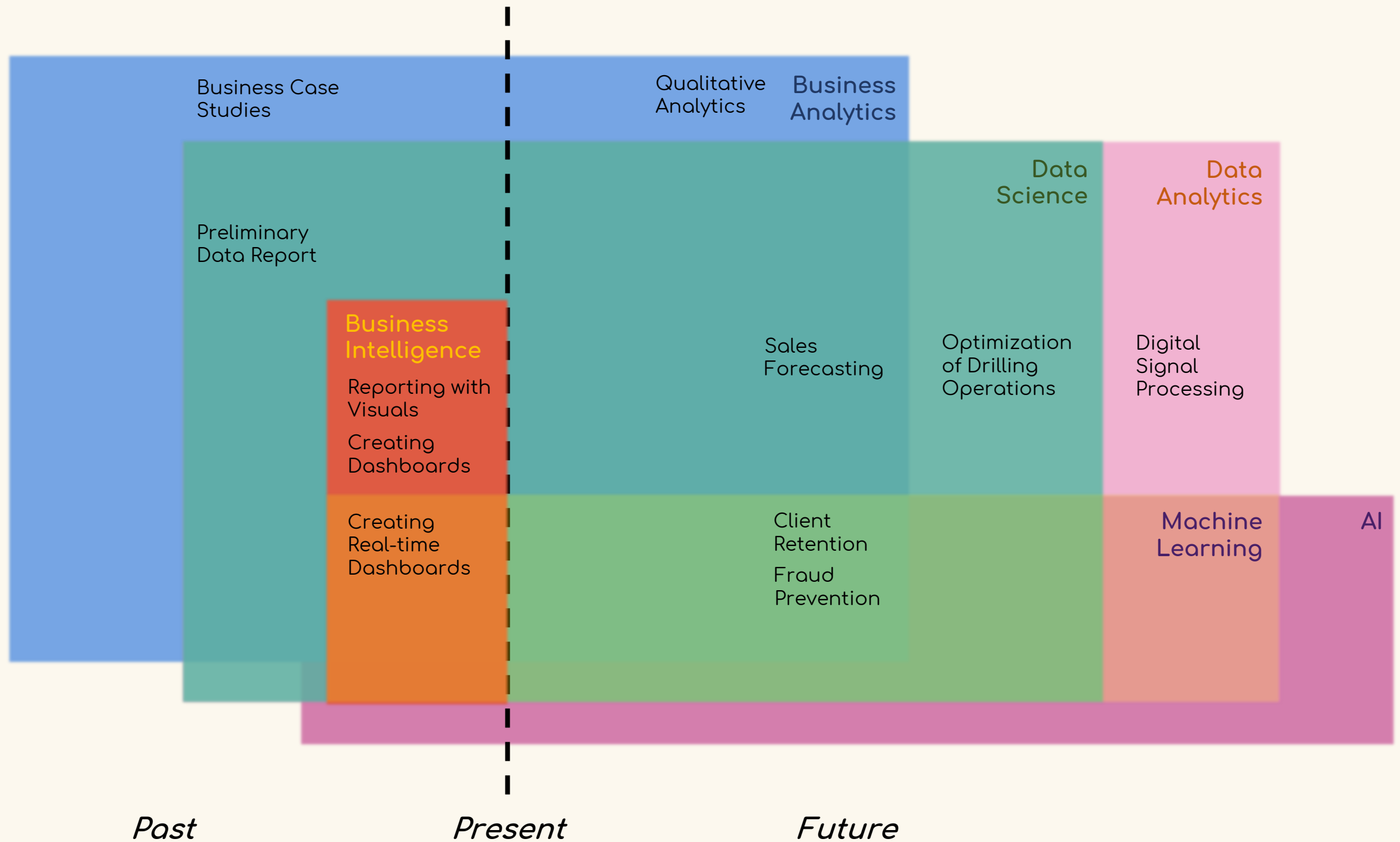
Machine
Learning

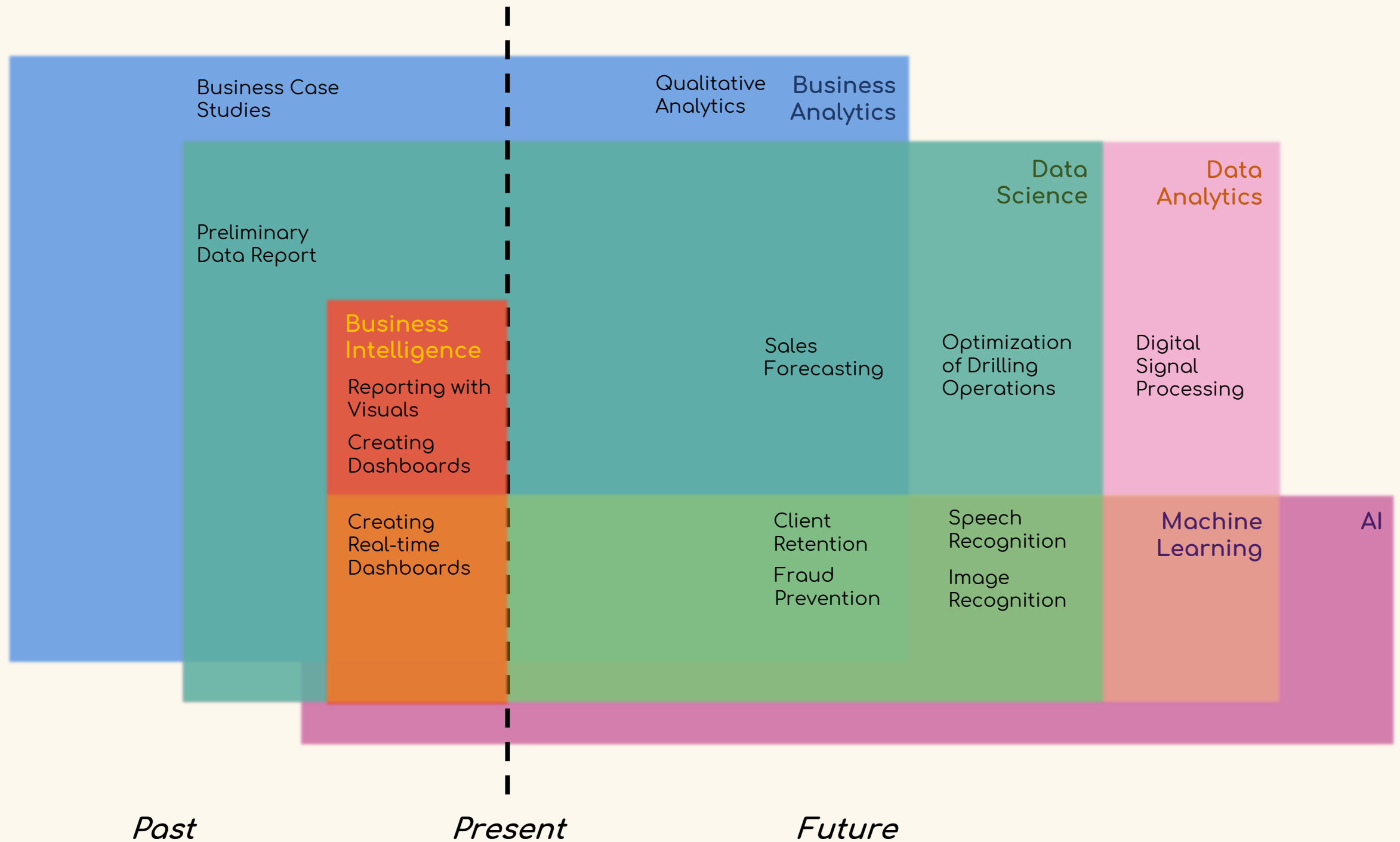
AI

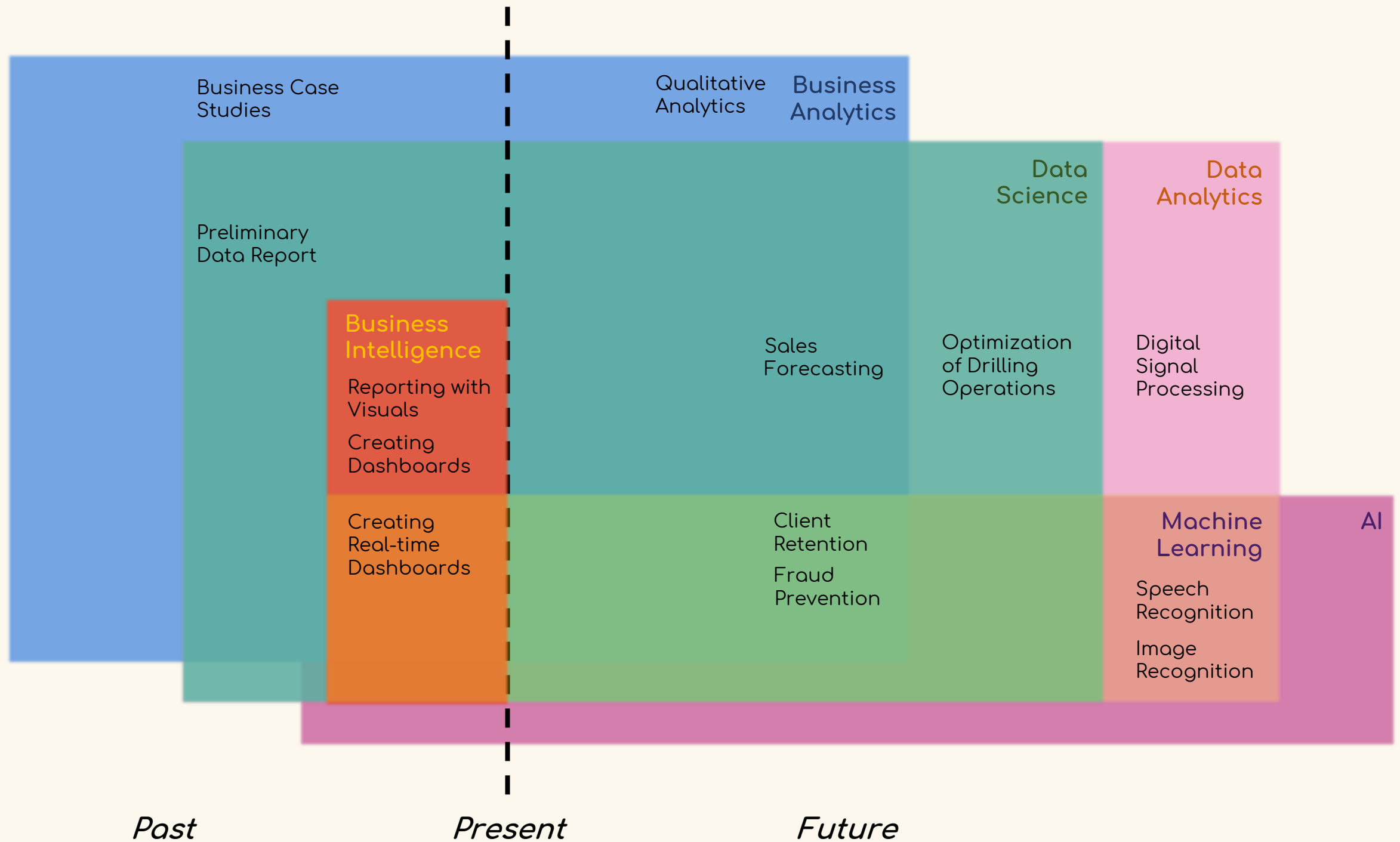
Past

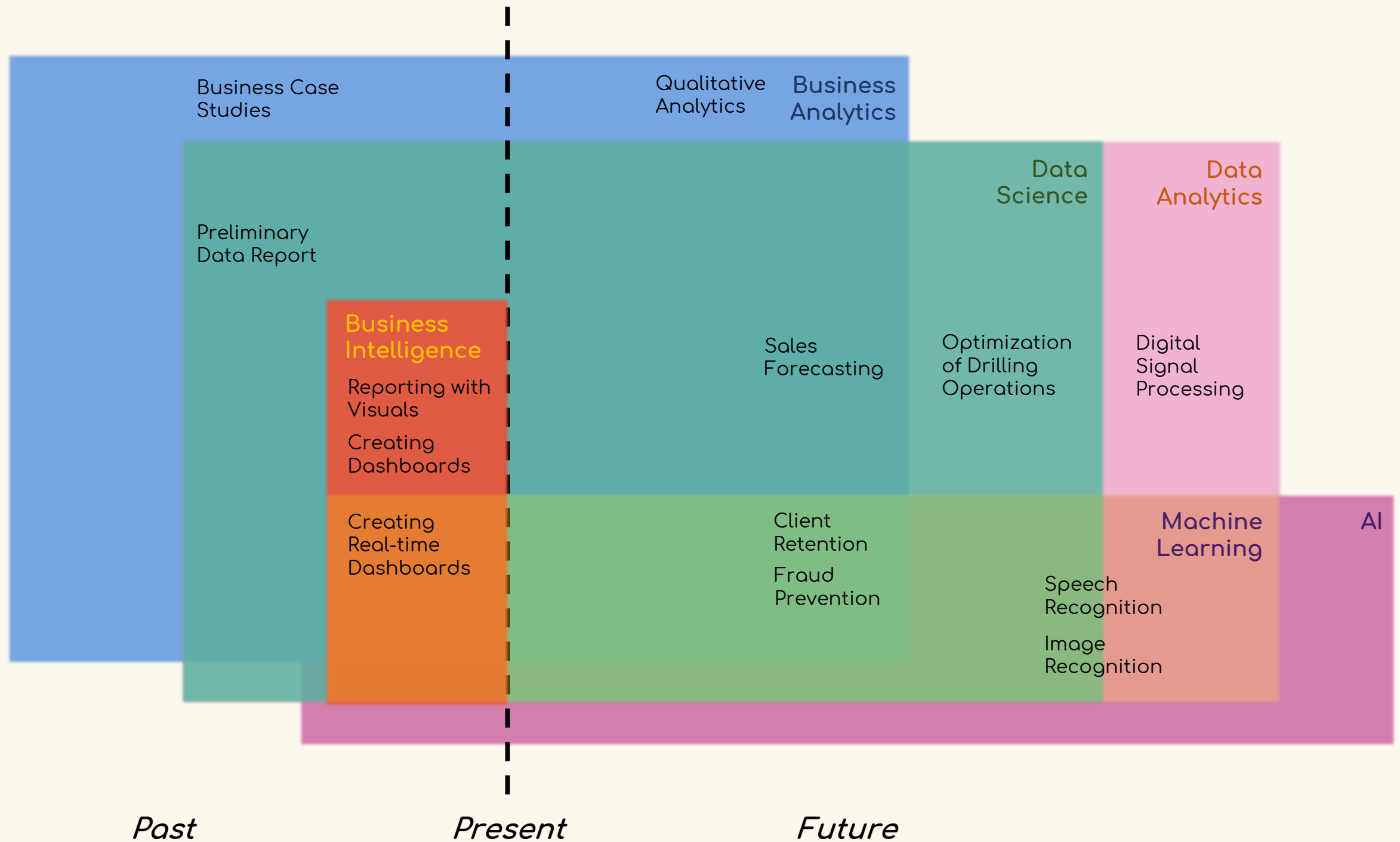
Present

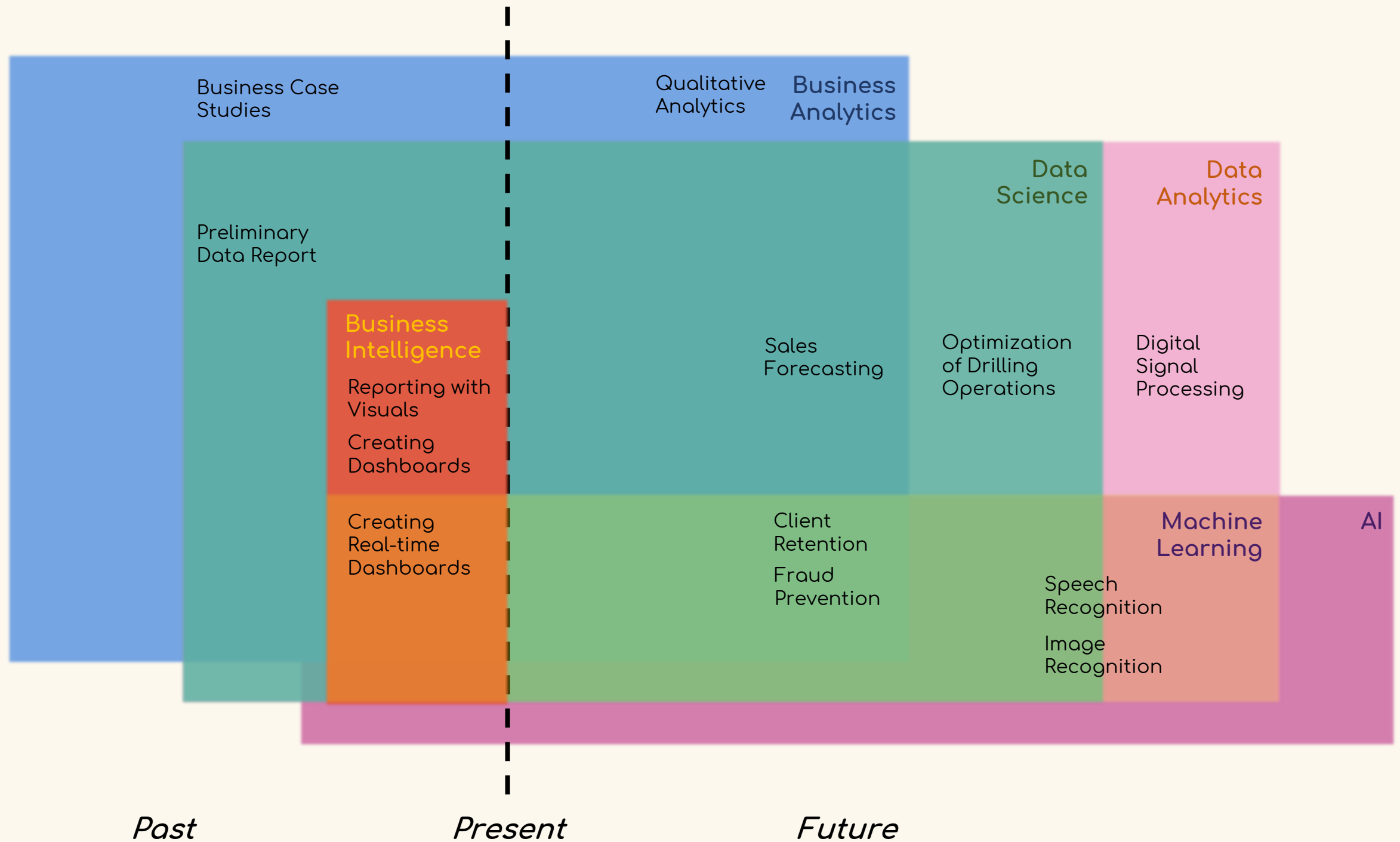
Future

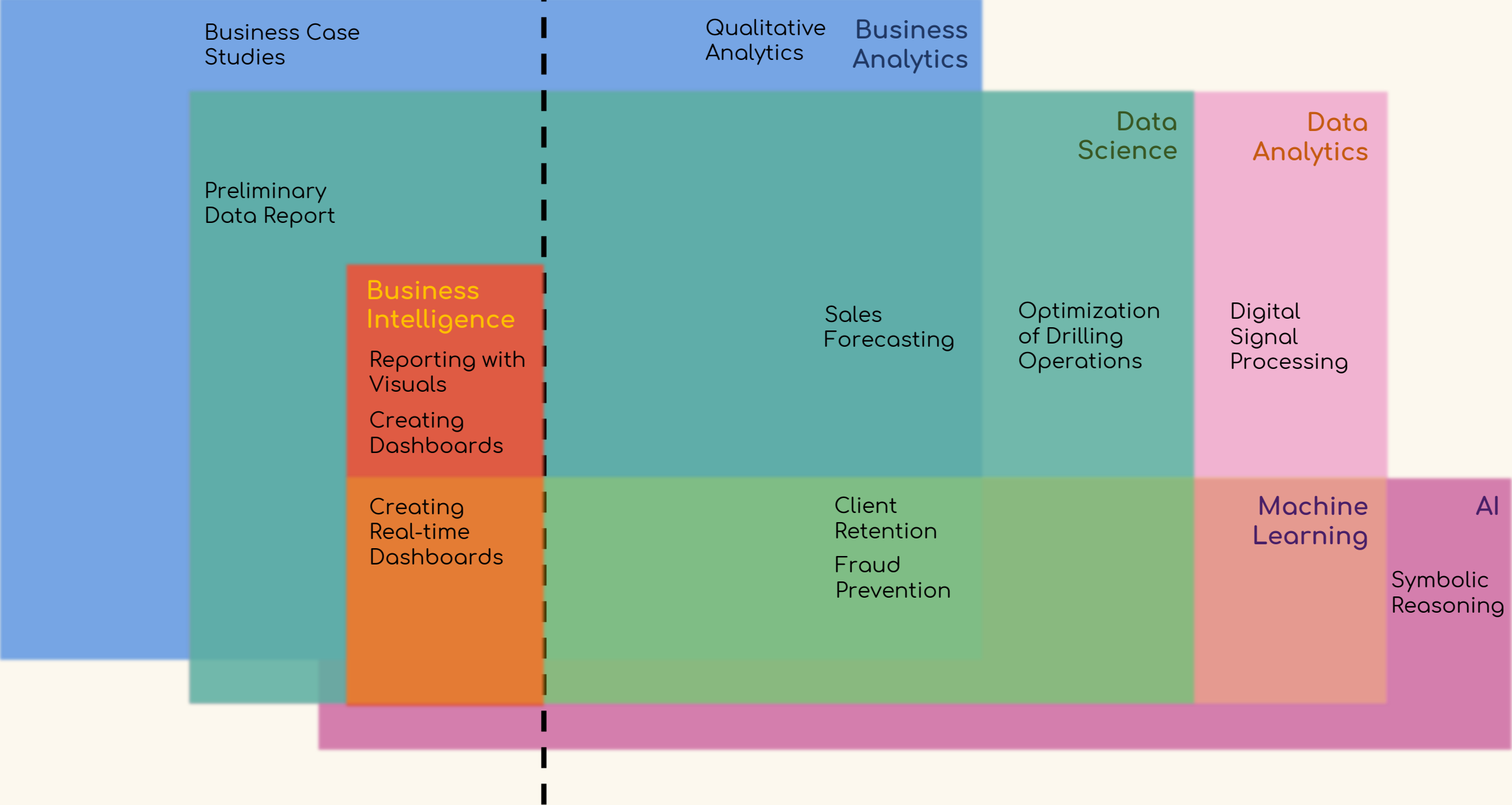








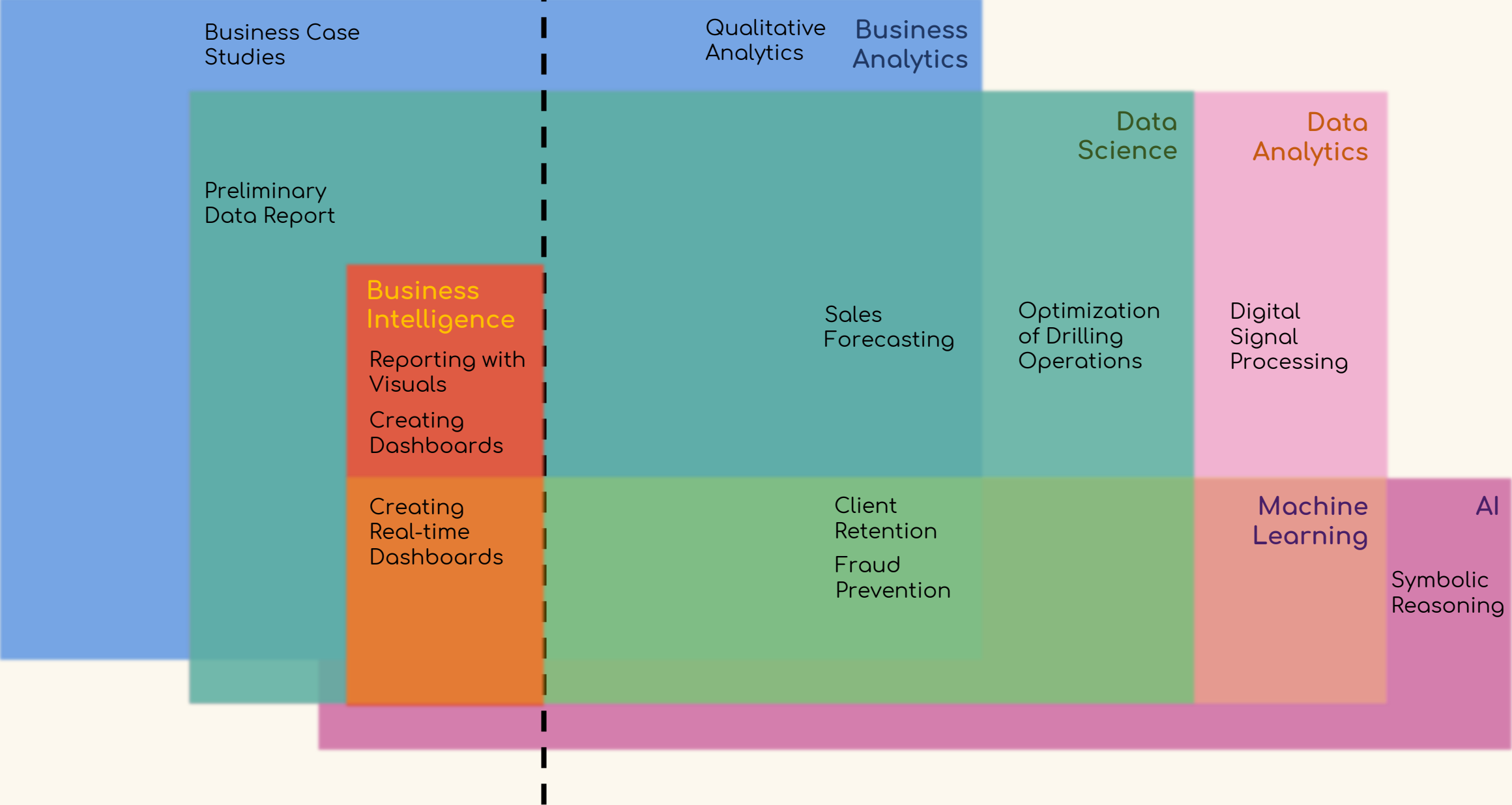




Past

Present

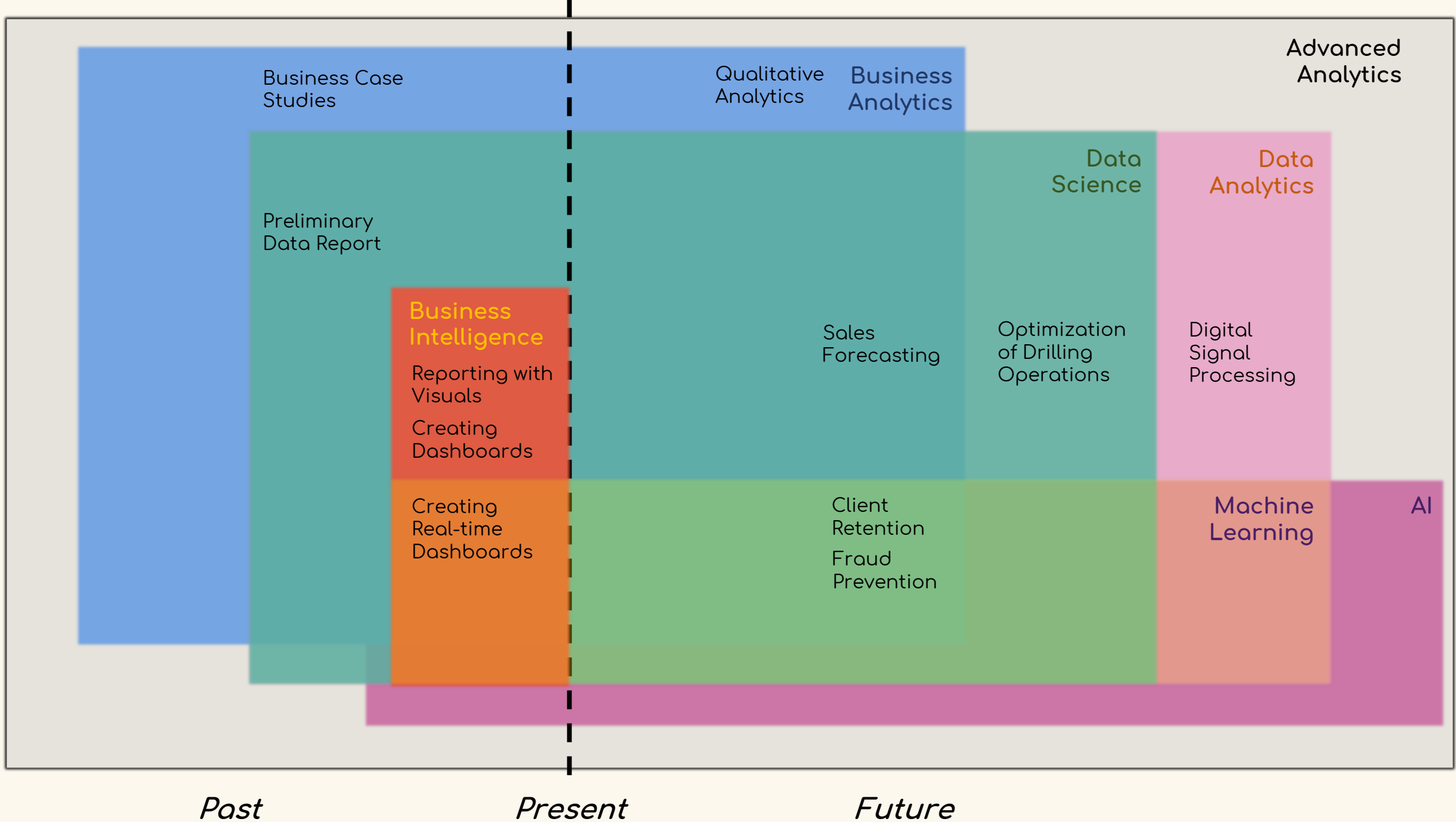
Future

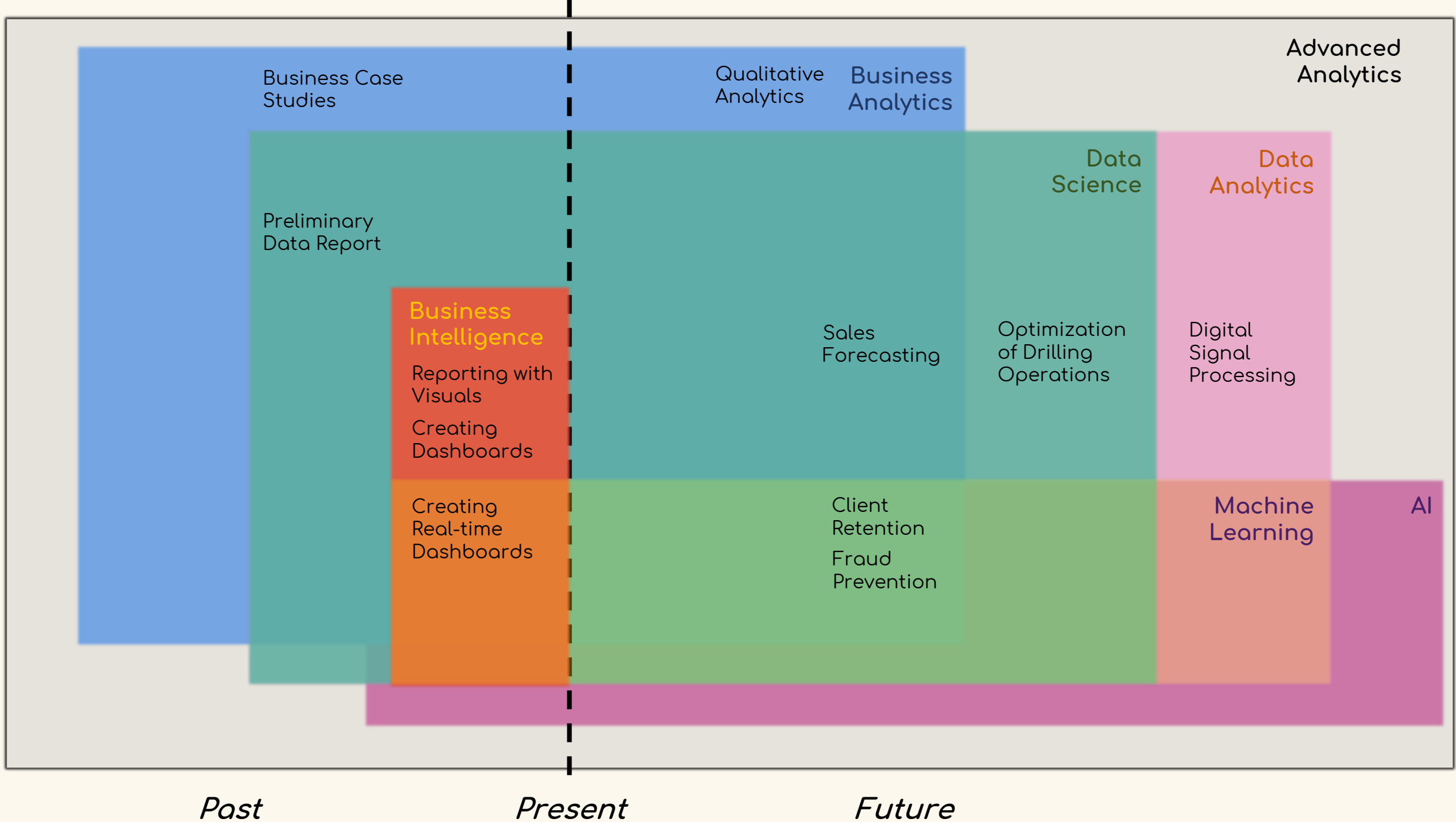


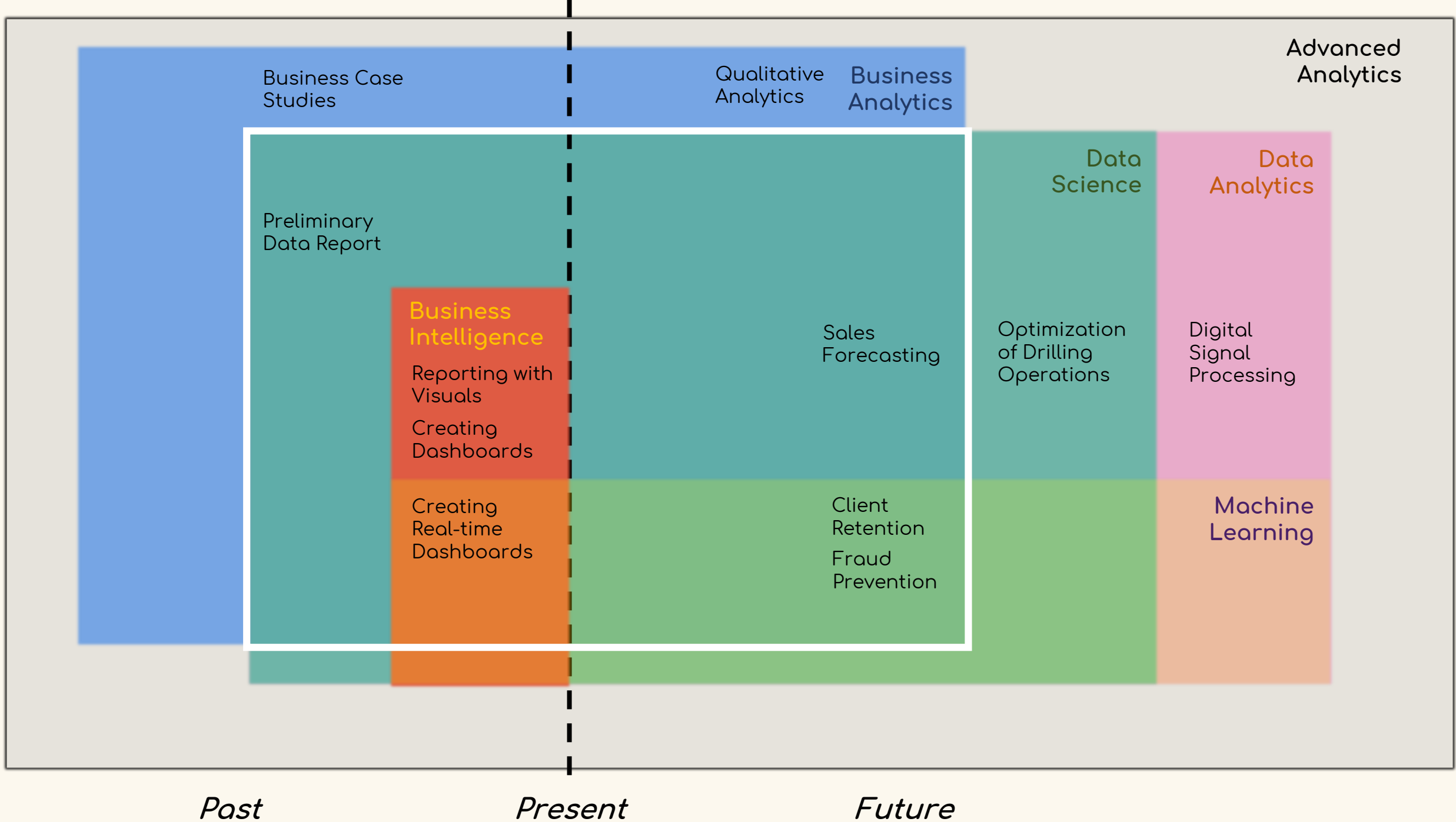
Past

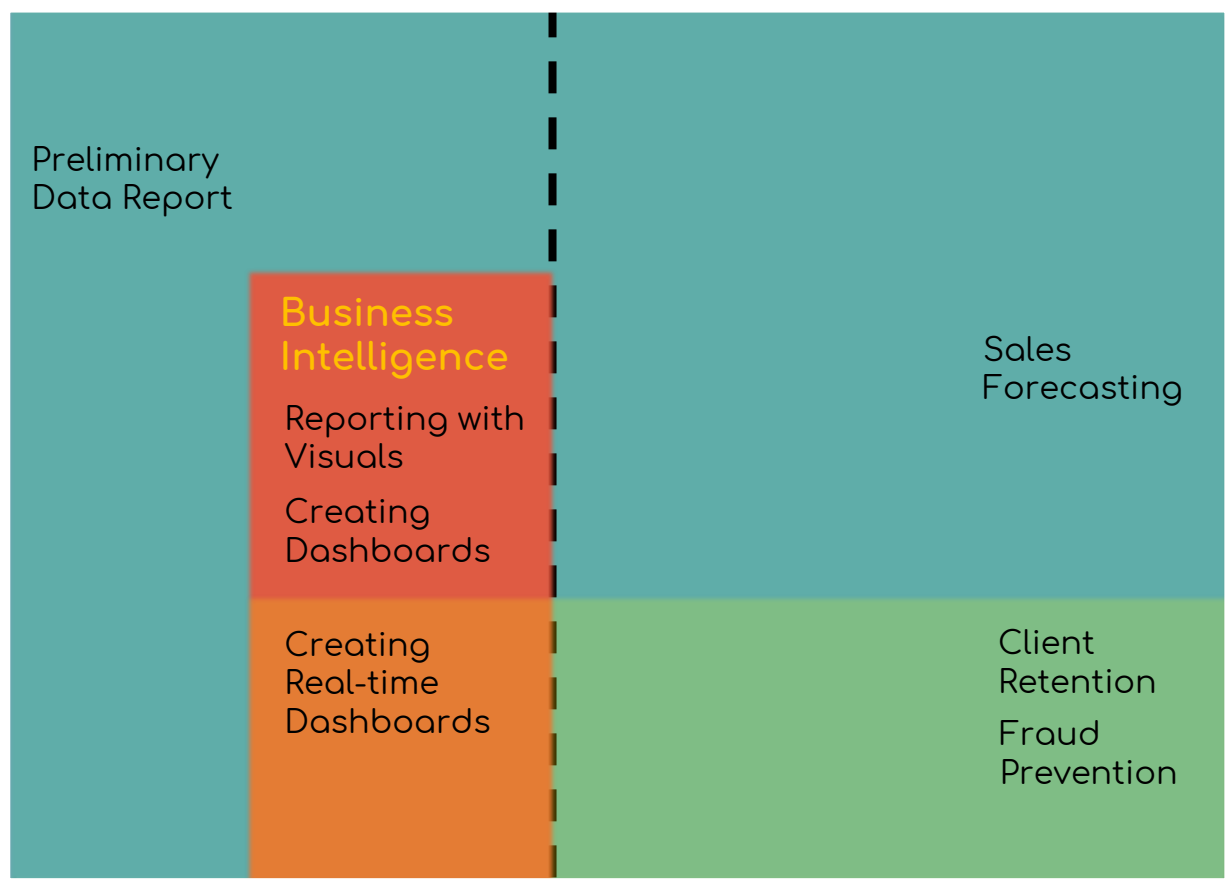
Present

Future





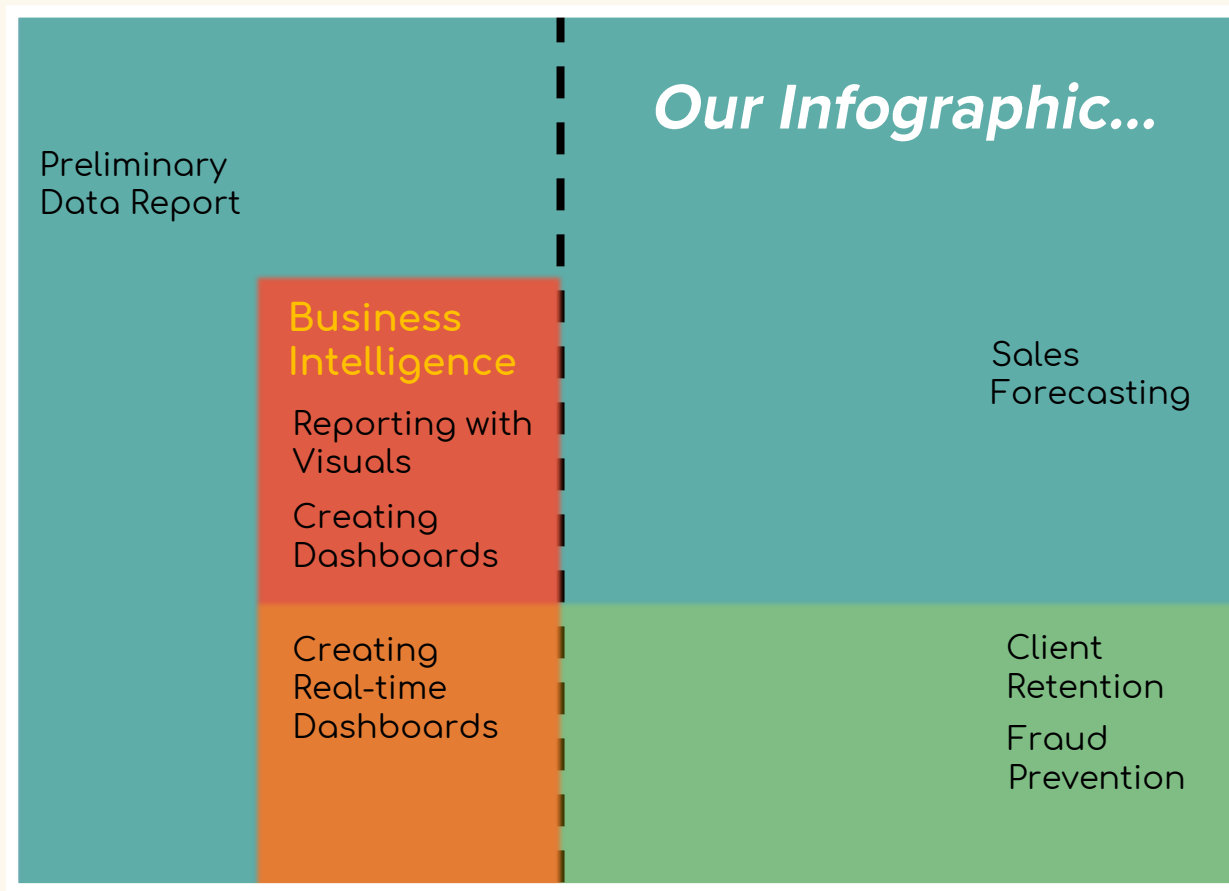




Past

Present

Future



Past

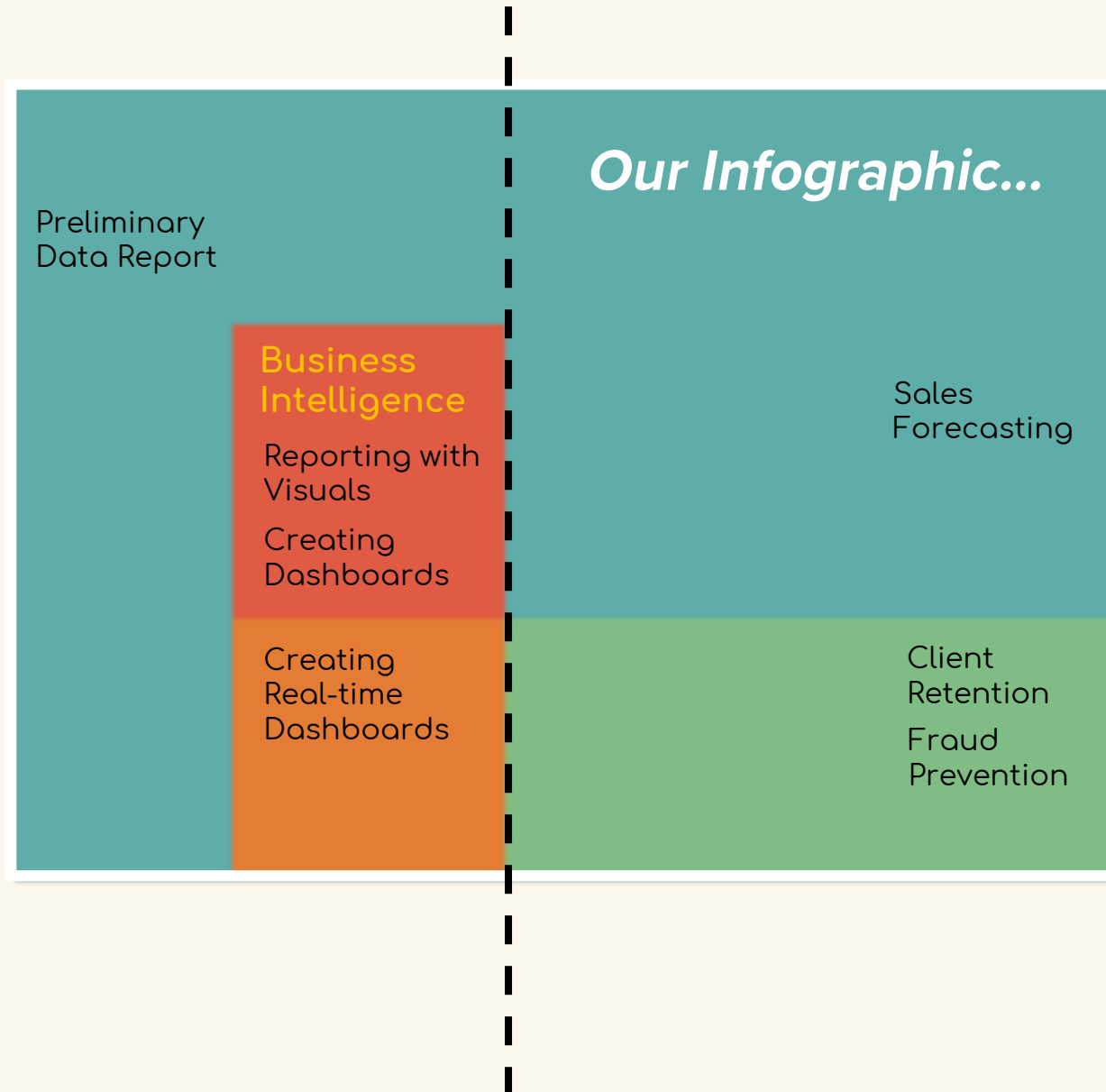
Present

Future

Past

Present

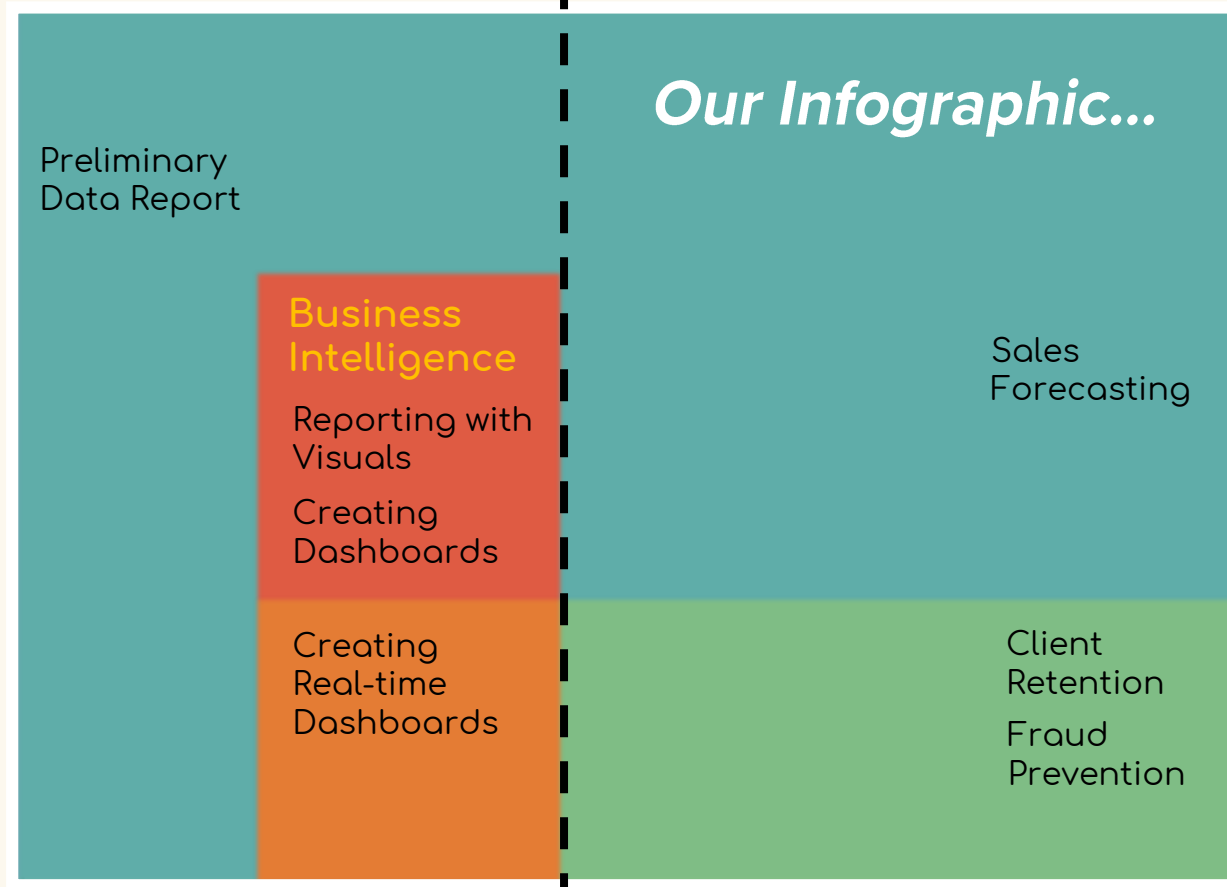
Future

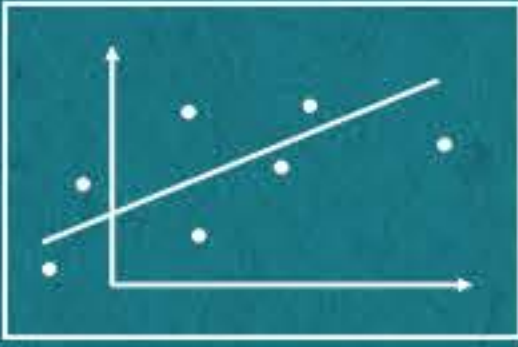
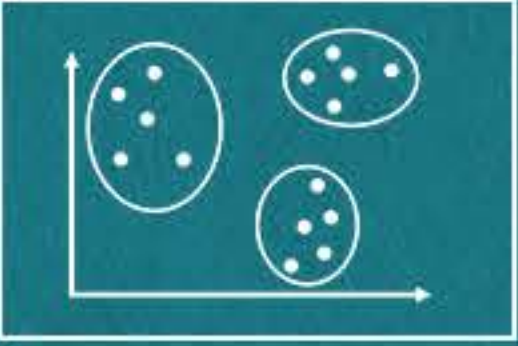

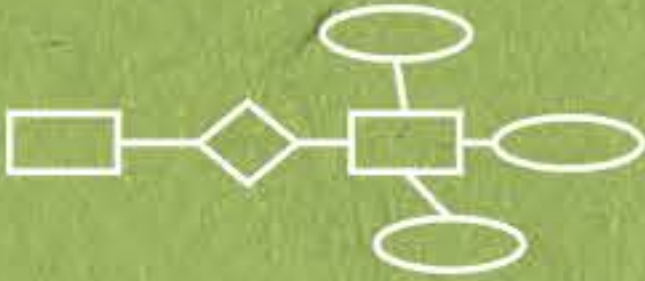
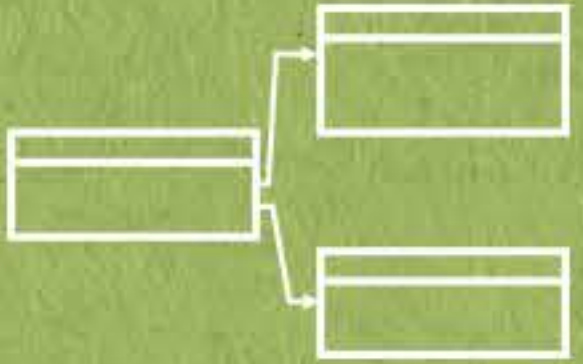
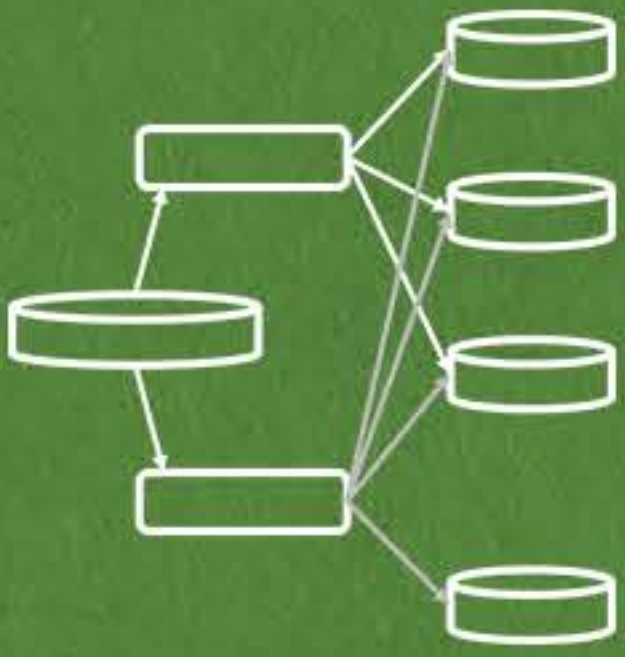













































Past

Present

Future



	Data		Data Science		
	TRADITIONAL	BIG	BUSINESS INTELLIGENCE	TRADITIONAL METHODS	MACHINE LEARNING
WHEN it is applied	At the beginning of your analysis		After the data has been gathered & organized	After BI reports have been created and discussed	
	PAST		NOW	FUTURE	
WHY you need it	data-driven decisions require well-organized and relevant raw data stored in a digital format		use data to create reports and dashboards to gain business insights	Predictive Analytics	
WHAT techniques are involved	DATA COLLECTION PREPROCESSING <ul style="list-style-type: none"> class labeling (categorical vs numerical) data cleansing dealing with missing values CASE SPECIFIC <ul style="list-style-type: none"> e.g. balancing & shuffling datasets 		ANALYZE THE DATA EXTRACT INFO AND PRESENT IT IN THE FORM OF: <ul style="list-style-type: none"> metrics KPIs reports dashboards 	REGRESSION  LOGISTIC REGRESSION CLUSTERING  FACTOR ANALYSIS TIME SERIES 	SUPERVISED LEARNING <ul style="list-style-type: none"> SVMs NNs deep learning random forests bayesian networks UNSUPERVISED LEARNING <ul style="list-style-type: none"> k-means deep learning REINFORCEMENT LEARNING similar to supervised learning, but instead of minimizing the loss, one maximizes reward
	  		   		
WHERE	BASIC CUSTOMER DATA HISTORICAL STOCK PRICE DATA	SOCIAL MEDIA FINANCIAL TRADING DATA	PRICE OPTIMIZATION INVENTORY MANAGEMENT	USER EXPERIENCE (UX) SALES FORECASTING	FRAUD DETECTION CLIENT RETENTION
HOW using what tools	PROGRAMMING LANGUAGES     SOFTWARE  	PROGRAMMING LANGUAGES     SOFTWARE   	PROGRAMMING LANGUAGES     SOFTWARE     	PROGRAMMING LANGUAGES    SOFTWARE    	PROGRAMMING LANGUAGES       SOFTWARE    
WHO	DATA ARCHITECT DATA ENGINEER DATABASE ADMINISTRATOR	BIG DATA ARCHITECT BIG DATA ENGINEER	BI ANALYST BI CONSULTANT BI DEVELOPER	DATA SCIENTIST DATA ANALYST	DATA SCIENTIST MACHINE LEARNING ENGINEER
ARE YOU AWARE	200,000 lines of data is not necessarily big data. It is not just volume that defines a data set as 'big' - variety, variability, velocity, veracity, and other characteristics are determinative as well.		Qualitative analysis tools such as SWOT are not used for quantitative analysis. Hence, they are not part of business intelligence.	Software like Excel, SPSS, and Stata, can be successfully used by data science teams in many companies.	In deep learning, there is still a debate on WHY the algorithms used outperform all conventional methods.



MACHINE LEARNING

INTRODUCTION

Machine learning

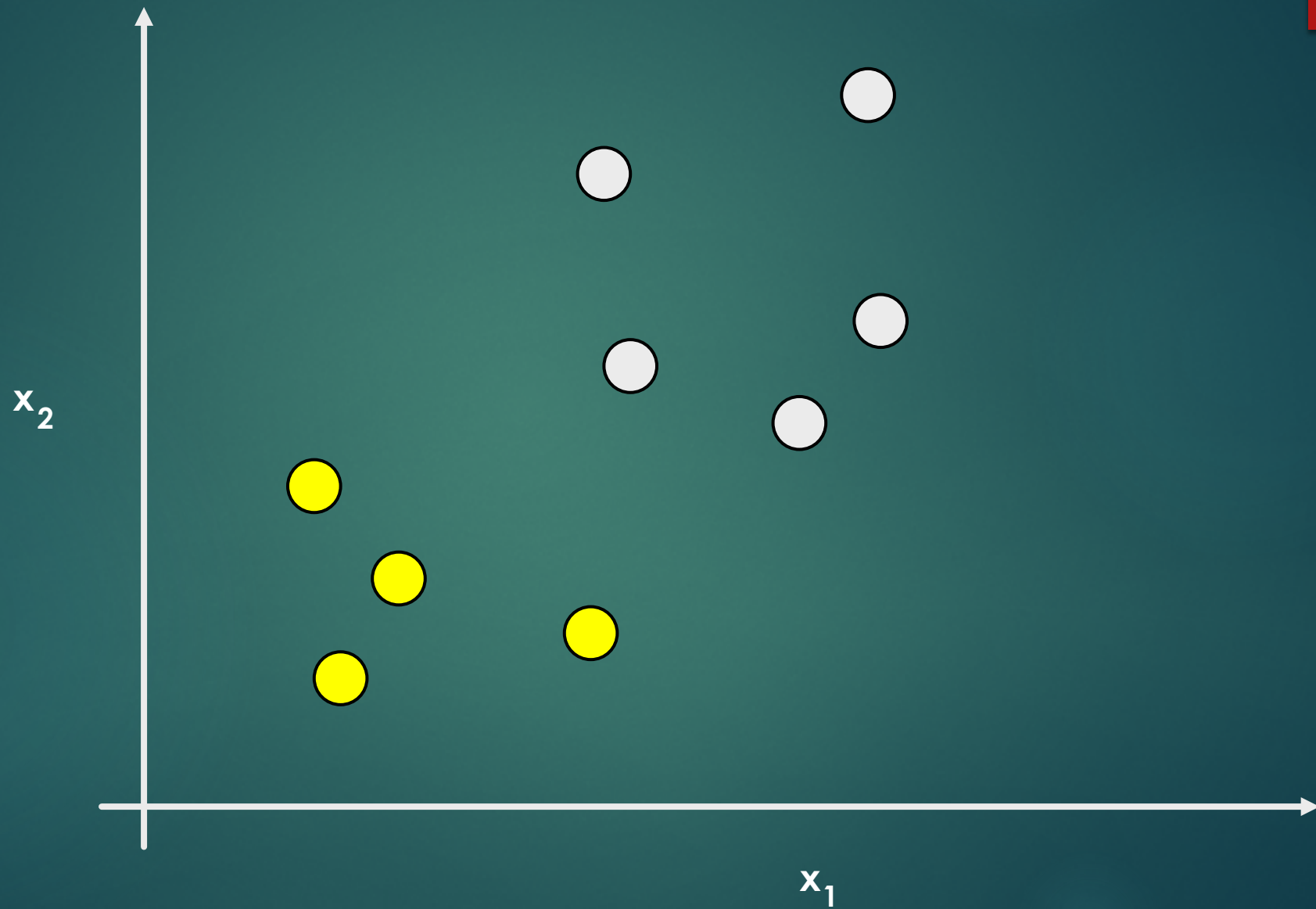
- ▶ This is a field of study that gives computers the ability to learn without being explicitly programmed
- ▶ Computer program performance on a given task improves with experience !!!
- ▶ Becoming more and more popular
- ▶ **Types: *supervised learning* AND *unsupervised learning***

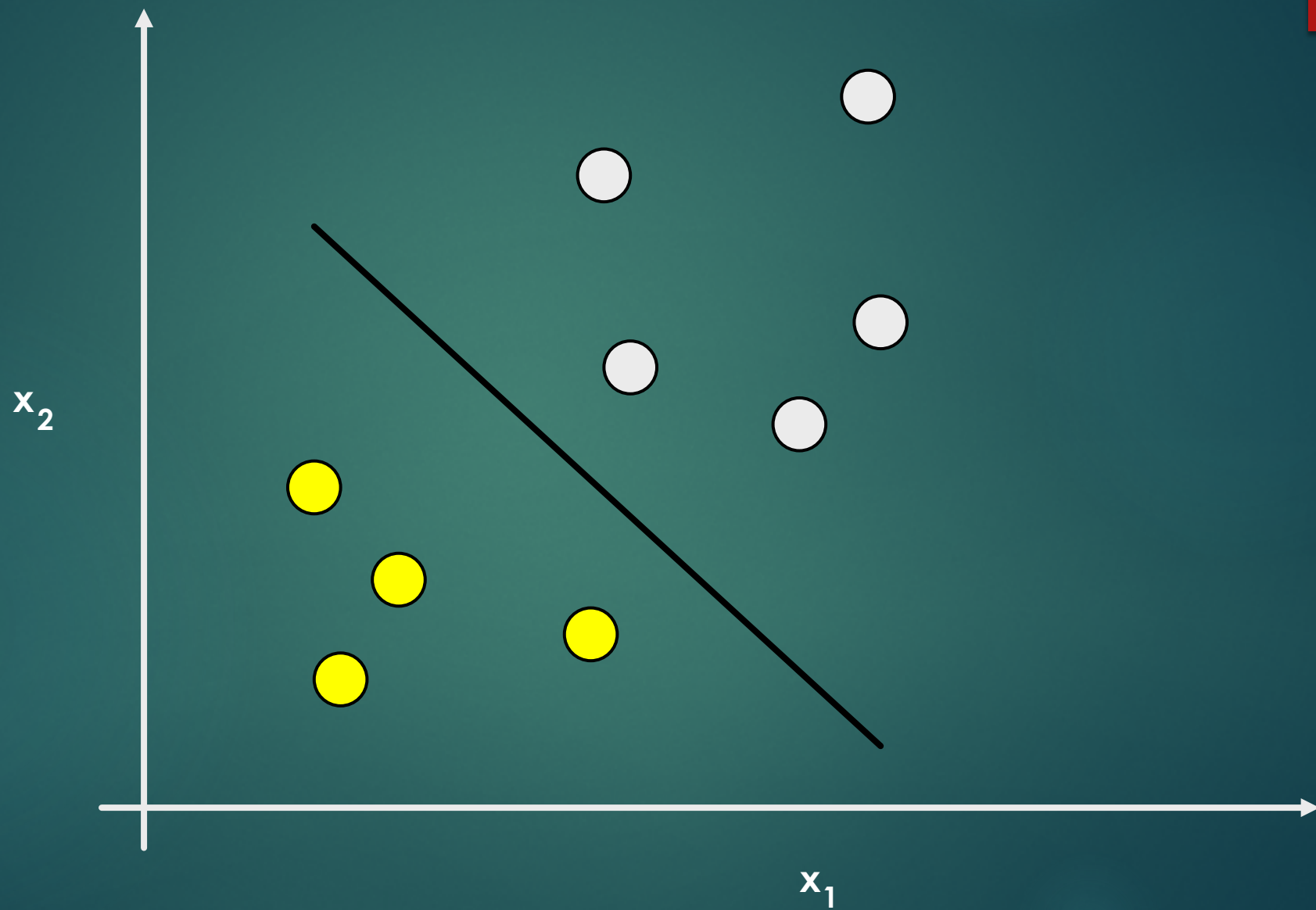
Supervised learning


- ▶ We give the algorithm the right answers
- ▶ For example: give prices of houses according to sizes (regression problems)
- ▶ Deciding whether a given email is a spam or not ... kind of a classification problem (logistic regression)
- ▶ Regression, naive Bayes, SVM, neural nets

AND logical relation

x	y	x AND y
0	0	0
0	1	0
1	0	0
1	1	1





- 
- ▶ **Important:** we have a dataset !!!
 - ▶ We usually split it up: training set and a test set // **80% - 20%**
 - ▶ We train our algorithm on the training set then we run our algorithm on the test set to check whether it is working fine or not
 - ▶ (we have two error terms: for the training set + for the test set)

Unsupervised learning

- ▶ We do not have a dataset with the right answers → we have just the data itself without labels !!!
- ▶ We have the data: can the algorithm find some pattern?
- ▶ The algorithm is able to find clusters
- ▶ For example: k-means clustering, hierarchical clustering, PCA
- ▶ Google news relies heavily on these techniques
- ▶ Genes classification
- ▶ Market segmentation or social networks

