

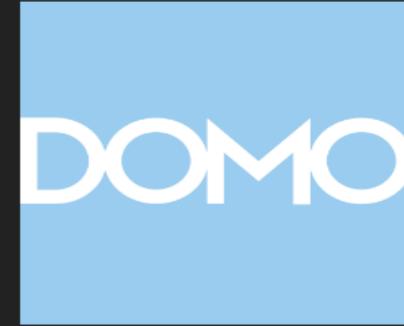
# Data Analysis Fundamentals

Data Science Career Overview

Satyam Govila

# Introduction to “Data”

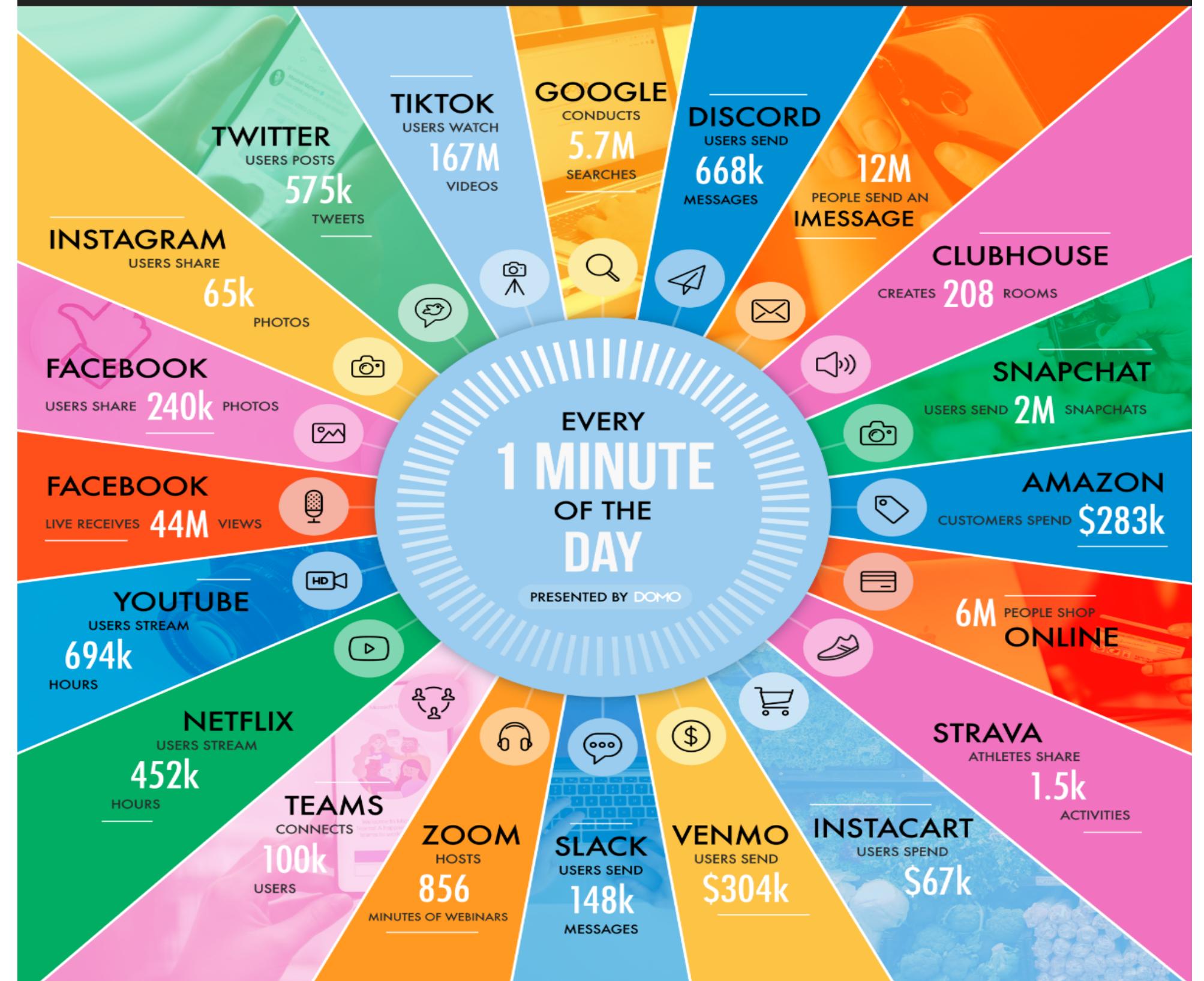
- Where does this data come from?
- How much data are we talking about?
- What are the types of data we are going to deal with?
- What to do with this data?



# Data Never Sleeps 9.0

## How much data is generated every minute?

The 2020 pandemic upended everything, from how we engage with each other to how we engage with brands and the digital world. At the same time, it transformed how we eat, how we work and how we entertain ourselves. Data never sleeps and it shows no signs of slowing down. In our 9th edition of the "Data Never Sleeps" infographic, we bring you a glimpse of how much data is created every digital minute in our increasingly data-driven world.



As of July 2021, the internet reaches 65% of the world's population and now represents 5.17 billion people—a 10% increase from January 2021. Of this total, 92.6 percent accessed the internet via mobile devices. According to Statista, the total amount of data consumed globally in 2021 was 79 zettabytes, an annual number projected to grow to over 180 zettabytes by 2025.

### Global Internet Population Growth (IN BILLIONS)



As the world changes, businesses need to change too—and that requires data. Domo gives you the power to make data-driven decisions at any moment, on any device, so that you can make smart choices in a rapidly changing world. Every click, swipe, share, or like tells you something about your customers and what they want, and Domo is here to help you and your business make sense of all of it.

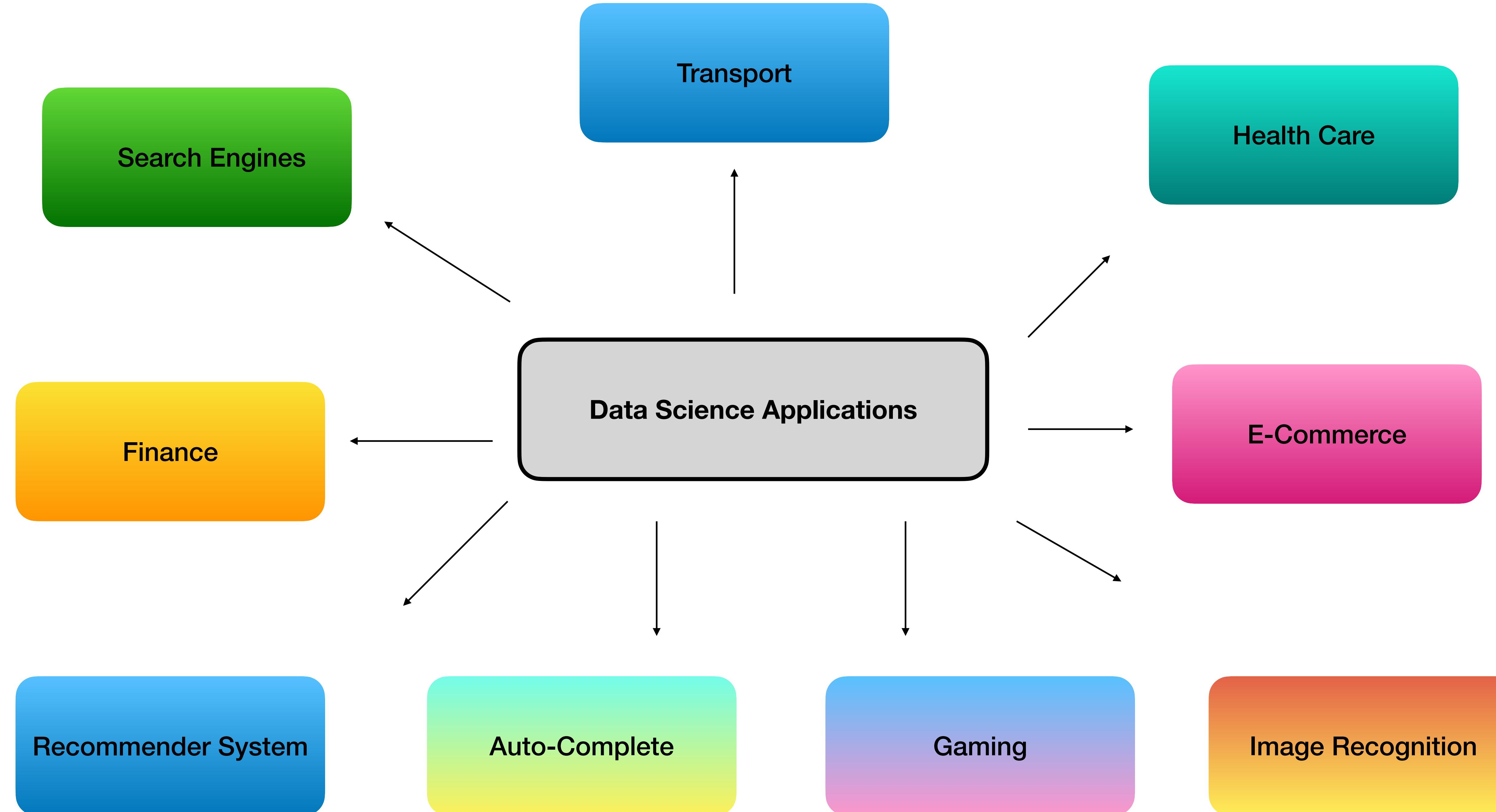
Learn more at [domo.com](https://domo.com)

SOURCES: LOCAL IQ, BUSINESS OF APPS, DUSTIN STOUT, HOOTSUITE, EXPANDED RAMBLINGS, INTERNET WORLD STATS, STATISTA, CNBC, BRANDWATCH, KILL THE CABLE BILL, YOUTUBE, KINSTA, THE VERGE, MANAGEMENT COMMUNICATION: A CASE ANALYSIS APPROACH, INTERNET LIVE STATS, SODA, STATISTA



# What is Data Science?

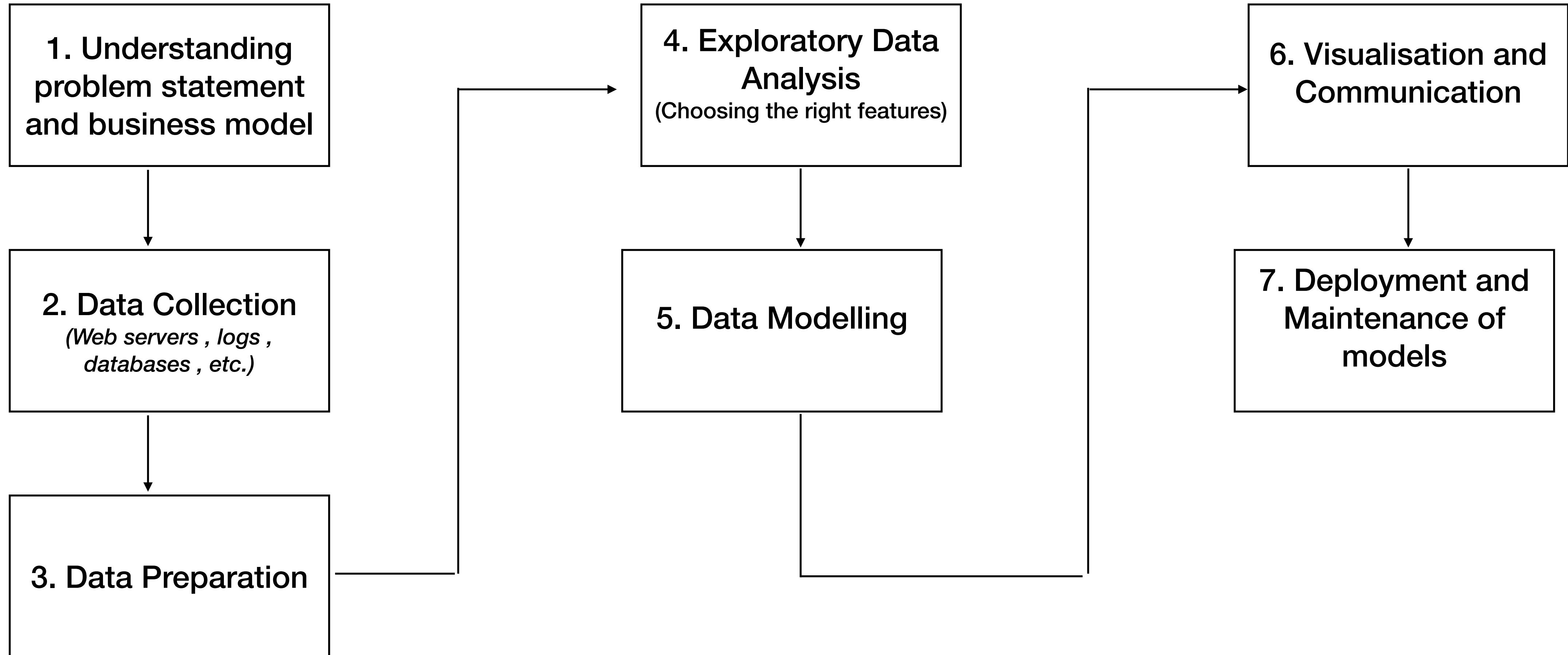
Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions!



# Data Science Lifecycle

**Case Study : Netflix Movie Recommendation System**

# Case Study : Netflix Movie Recommendation System



The diagram illustrates three types of data quality issues using two tables:

- Improper Datatype:** Indicated by an orange arrow pointing to the first column of the left table, which contains a mix of numerical and categorical values.
- Missing Value:** Indicated by an orange arrow pointing to the cell at row 5, column 2 (Carats: 0.37, Price: null).
- Null Value:** Indicated by an orange arrow pointing to the cell at row 7, column 2 (Carats: 0.63, Price: null).

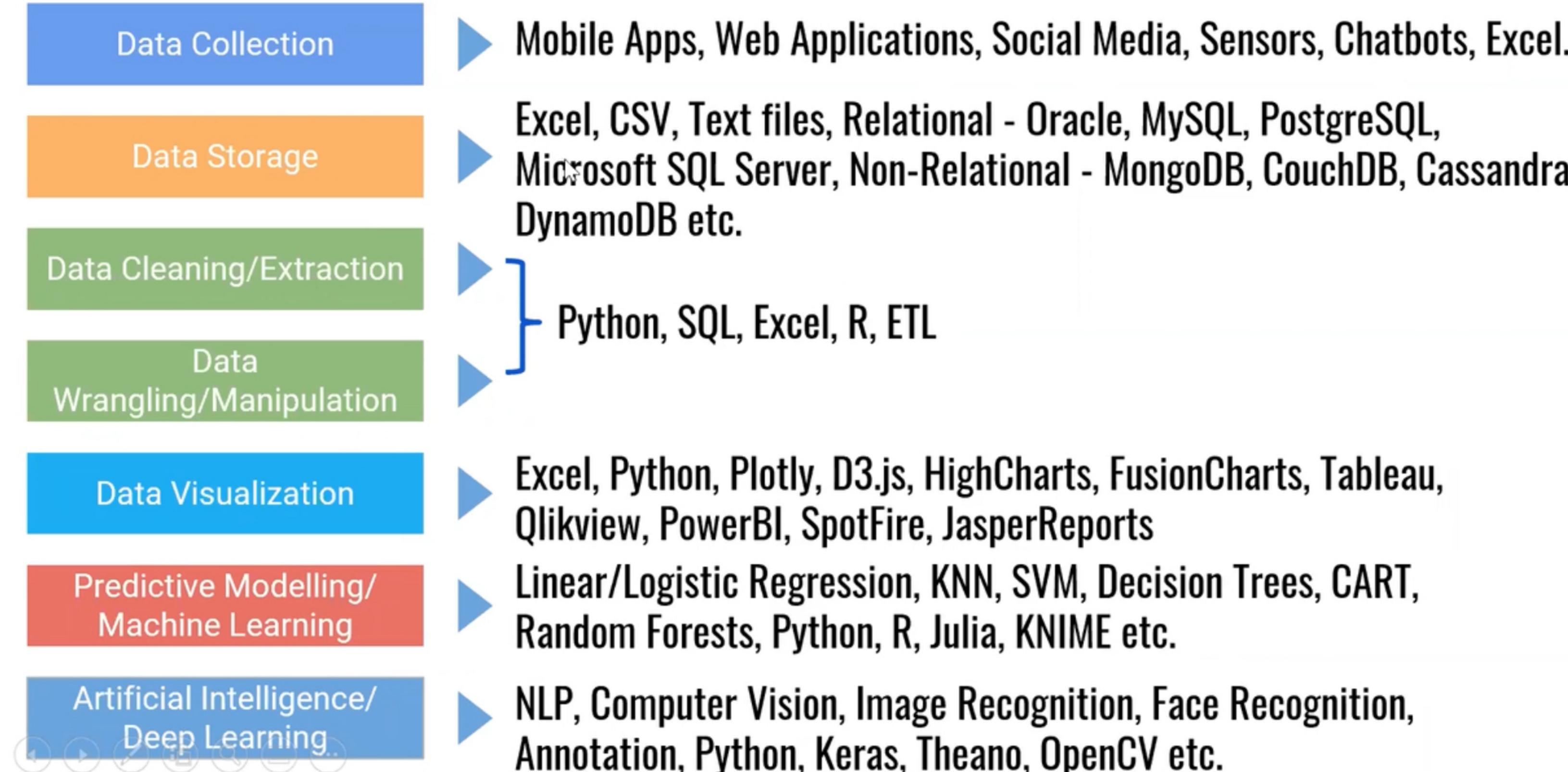
**Left Table (Initial Data):**

B	C
Carats	Price
1.01	7366
0.49	985
0.31	544
1.51	140
0.37	
0.73	3011
1.53	11413
0.56	1814
0.41	876
0.74	2690
0.63	NULL
0.6	4172
Two	11764
1.1	4682
1.31	6171

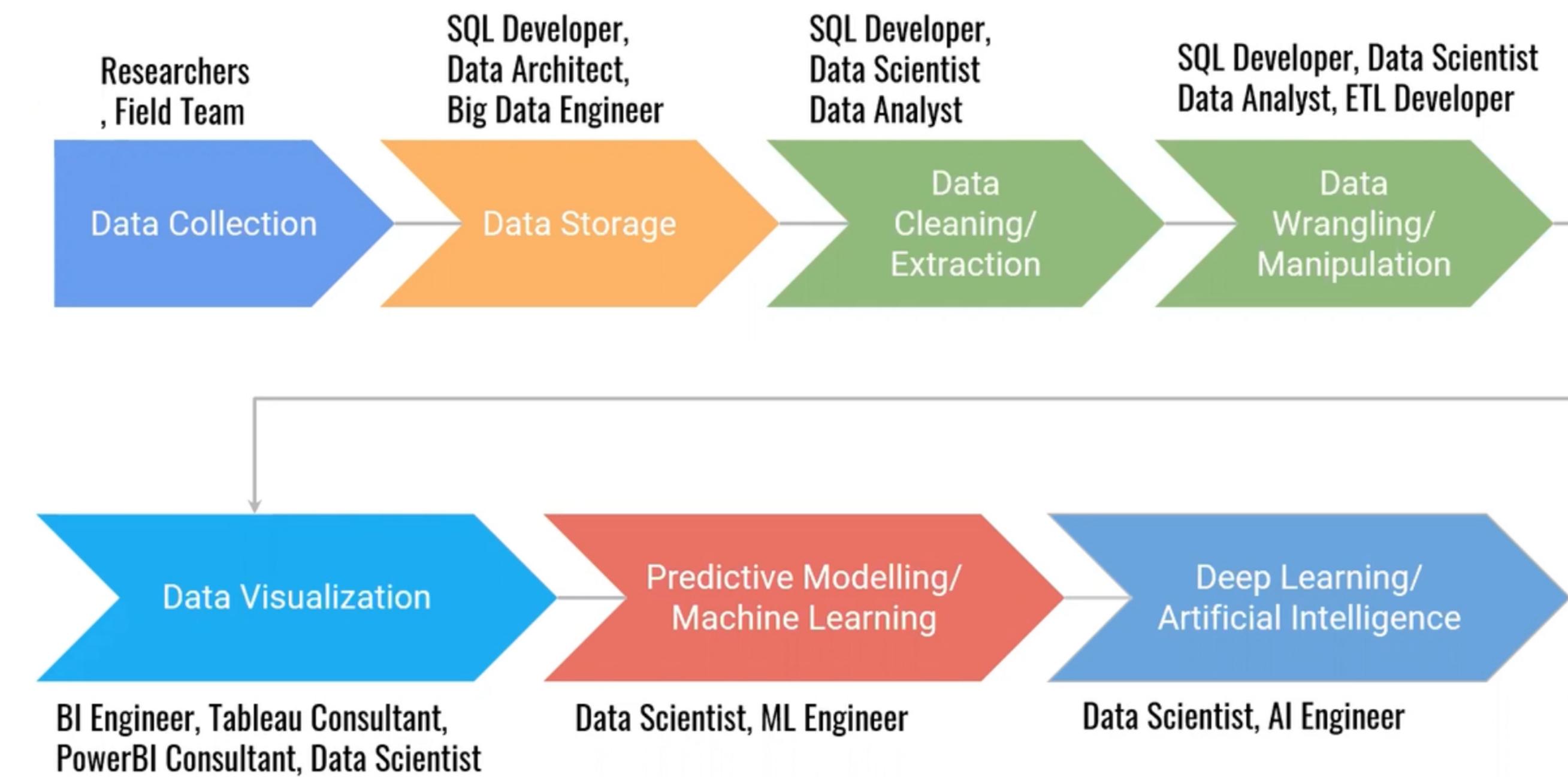
**Right Table (Cleaned Data):**

B	C
Carats	Price
1.01	7366
0.49	985
0.31	544
1.51	140
0.37	493
0.73	3011
1.53	11413
0.56	1814
0.41	876
0.74	2690
0.63	1190
0.6	4172
2	11764
1.1	4682
1.31	6171

# Building Blocks of Data Science



# Different Roles in Data Science



# Any Questions?