Topics in Computer Science

Lecture 3

Melvin Ayala

Lecture 3: Data Science Analytics and Methodology

Sections:

1. Data Analytics in Practice

Decision support:

- Original reason for development of:
  - Big data
  - Business intelligence (BI)
  - Analytics
  - Performance management,
  - Data lakes
  - AI

- However, usage has been expanded to other areas other than decision support

1.1. Data Analytics Lifecycle

Data Analytics Lifecycle:

- composed of 5 stages
- allows driving business value and innovation through continuous improvement.

Stage #	Name	Question	Tasks
1	PLANNING ANALYTICS	**What is your plan?**	Understands the status of your business.
2	DESCRIPTIVE ANALYTICS	**What happened?**	Identify what happened and determine how you are performing against the plan.
3	DIAGNOSTIC ANALYTICS	**Why did it happen?**	Answer this question and then identify key patterns.
4	PREDICTIVE ANALYTICS	**What happens next?**	Use those patterns to predict future trends.
5	PRESCRIPTIVE ANALYTICS	**What should you do next?**	Finally, use prescriptive analytics to decide what you should do to overcome operational obstacles found in your analysis.

1.2. Planning Analytics

Chicago-based Bikeshare Company Case Study

Chicago-based bikeshare company offers bike ride services around the globe.

Rides in October 2017:
42,567

Rides in 2017:
563,567

Total lifetime rides (2014-2017):
6,449,778

Process Details

Business Understanding

1	PLANNING ANALYTICS	What is your plan?
2	DESCRIPTIVE ANALYTICS	What happened?
3	DIAGNOSTIC ANALYTICS	Why did it happen?
4	PREDICTIVE ANALYTICS	What happens next?
5	PRESCRIPTIVE ANALYTICS	What should you do next?

Question: What is your Plan?

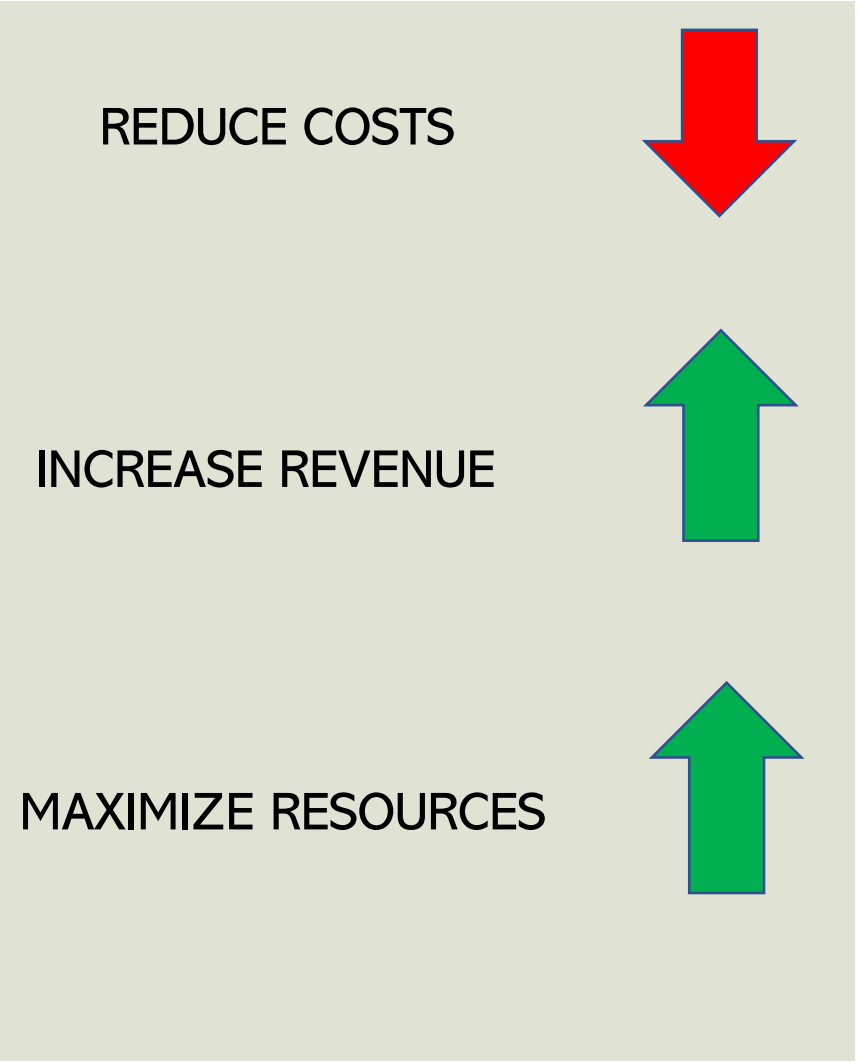
- Planning relies on outputs of all the other steps.
- It requires an understanding of past performance
- Identification of deviations from the norm (plan vs. actual)
- Evaluation of possible scenarios
- Prediction of likely outcomes, and assessment of risks and constrains.

Case Study: Bikeshare

Business Understanding

Question: What is your Plan?

- Understand budget constraints, revenue forecast, and provide ad-hoc analysis in real time.
- Perform demand planning and supply chain optimization for bikes across different segments
- Forecast operation spending and maximize capital allocations to meet revenue objectives.



Process Details

What is your plan? **Business Understanding**

Chicago Bikeshare

	Q1 2017			Q2 2017			Q3 2017	
Section	Actual	Budget	Variance %	Actual	Budget	Variance %	Budget	Variance --
Capital Grants and Contributions	90,602	99,073	-8.55%	104,873	99,073	5.85%	99,073	99,073
Gifts in Kind - Goods	22,494	20,271	10.96%	22,260	20,271	9.81%	20,271	20,271
Operating Grants and Contributions	79,819	79,360	0.58%	82,660	70,360	4.15%	79,300	79,300
Sponsorships	222,224	236,713	-5.72%	226,502	236,713	-3.91%	236,713	235,713
Membership Fees	636,300	681,172	-6.59%	1,301,155	1,442,608	-9.81%	1,844,668	1,844,668
Usage Fees	313,328	338,225	-7.30%	1,831,190	2,079,988	-11.96%	3,099,877	3,099,877
Membership and Daily Use Revenues	949,628	1,019,397	-6.84%	3,132,345	3,522,596	-11.06%	4,944,545	4,944,545
Accounts Receivable	63,384	61,027	-3.86%	66,519	61,027	-9.00%	61,027	61,027
Total Operating Resources	1,337,549	1,415,768	-5.52%	3,530,278	3,918,967	-9.92%	5,340,918	5,340,918
Total Income	1,426,151	1,514,841	-5.72%	3,635,149	4,016,040	-9.53%	5,439,969	5,439,969
Total Operating Expenses	328,583	326,263	0.71%	298,263	326,253	-8.58%	326,253	326,253
Liability Insurance	67,702	65,849	2.81%	66,779	65,849	1.41%	65,849	65,849
Marketting	8,972	8,862	1.23%	8,064	8,862	-9.13%	8,862	8,862
Office Expenses	67,645	50,285	14.64%	52,431	50,285	4.27%	50,285	50,285
Professional Services	13,820	12,563	10.01%	13,716	12,563	9.17%	12,563	12,563
Other Gen and Admin Expenses	40,181	39,588	1.50%	39,744	39,588	0.39%	39,588	39,588
G&A Payrol Expenses	16,428	16,304	7.35%	16,830	16,304	9.97%	15,304	15,304
Total General and Administrative	204,748	192,452	6.39%	197,663	192,452	2.65%	192,452	192,452
Total Expenses	633,332	518,704	2.82%	495,815	518,704	-4.41%	516,704	518,704

Process Details

What is your plan? **Business Understanding**

Chicago Bikeshare

	Q1 2017			Q2 2017			Q3 2017	
Section	Actual	Budget	Variance %	Actual	Budget	Variance %	Budget	Variance --
Capital Grants and Contributions	90,602	99,073	-8.55%	104,873	99,073	5.85%	99,073	99,073
Gifts in Kind - Goods	22,494	20,271	10.96%	22,260	20,271	9.81%	20,271	20,271
Operating Grants and Contributions	79,819	79,360	0.58%	82,660	70,360	4.15%	79,300	79,300
Sponsorships	222,224	236,713	-5.72%	226,502	236,713	-3.91%	236,713	235,713
Membership Fees	636,300	681,172	-6.59%	1,301,155	1,442,608	-9.81%	1,844,668	1,844,668
Usage Fees	313,328	338,225	-7.30%	1,831,190	2,079,988	-11.96%	3,099,877	3,099,877
Membership and Daily Use Revenues	949,628	1,019,397	-6.84%	3,132,345	3,522,596	-11.08%	4,944,545	4,944,545
Accounts Receivable	63,384	61,027	-3.86%	66,519	61,027	-9.00%	61,027	61,027
Total Operating Resources	1,337,549	1,415,768	-5.52%	3,530,278	3,918,967	-9.92%	5,340,918	5,340,918
Total Income	1,426,151	1,514,841	-5.72%	3,635,149	4,016,040	-9.53%	5,439,969	5,439,969
Total Operating Expenses	328,583	326,263	0.71%	298,263	326,253	-8.58%	326,253	326,253
Liability Insurance	67,702	65,849	2.81%	66,779	65,849	1.41%	65,849	65,849
Marketting	8,972	8,862	1.23%	8,064	8,862	-9.13%	8,862	8,862
Office Expenses	67,645	50,285	14.64%	52,431	50,285	4.27%	50,285	50,285
Professional Services	13,820	12,563	10.01%	13,716	12,563	9.17%	12,563	12,563
Other Gen and Admin Expenses	40,181	39,588	1.50%	39,744	39,588	0.39%	39,588	39,588
G&A Payrol Expenses	16,428	16,304	7.35%	16,830	16,304	9.97%	15,304	15,304
Total General and Administrative	204,748	192,452	6.39%	197,663	192,452	2.65%	192,452	192,452
Total Expenses	633,332	518,704	2.82%	495,815	518,704	-4.41%	516,704	518,704

Business Insights:

Q2 revenue targets were off by 11.08%

Process Details

What is your plan? Business Understanding

Chicago Bikeshare

	Q1 2017			Q2 2017			Q3 2017	
Section	Actual	Budget	Variance %	Actual	Budget	Variance %	Budget	Variance --
Capital Grants and Contributions	90,602	99,073	-8.55%	104,873	99,073	5.85%	99,073	99,073
Gifts in Kind - Goods	22,494	20,271	10.96%	22,260	20,271	9.81%	20,271	20,271
Operating Grants and Contributions	79,819	79,360	0.58%	82,660	70,360	4.15%	79,300	79,300
Sponsorships	222,224	236,713	-5.72%	226,502	236,713	-3.91%	236,713	235,713
Membership Fees	636,300	681,172	-6.59%	1,301,155	1,442,608	-9.81%	1,844,668	1,844,668
Usage Fees	313,328	338,225	-7.30%	1,831,190	2,079,988	-11.96%	3,099,877	3,099,877
Membership and Daily Use Revenues	949,628	1,019,397	-6.84%	3,132,345	3,522,596	-11.08%	4,944,545	4,944,545
Accounts Receivable	63,384	61,027	-3.86%	66,519	61,027	-9.00%	61,027	61,027
Total Operating Resources	1,337,549	1,415,768	-5.52%	3,530,278	3,918,967	-9.92%	5,340,918	5,340,918
Total Income	1,426,151	1,514,841	-5.72%	3,635,149	4,016,040	-9.53%	5,439,969	5,439,969
Total Operating Expenses	328,583	326,263	0.71%	298,263	326,253	-8.58%	326,253	326,253
Liability Insurance	67,702	65,849	2.81%	66,779	65,849	1.41%	65,849	65,849
Marketting	8,972	8,862	1.23%	8,064	8,862	-9.13%	8,862	8,862
Office Expenses	67,645	50,285	14.64%	52,431	50,285	4.27%	50,285	50,285
Professional Services	13,820	12,563	10.01%	13,716	12,563	9.17%	12,563	12,563
Other Gen and Admin Expenses	40,181	39,588	1.50%	39,744	39,588	0.39%	39,588	39,588
G&A Payrol Expenses	16,428	16,304	7.35%	16,830	16,304	9.97%	15,304	15,304
Total General and Administrative	204,748	192,452	6.39%	197,663	192,452	2.65%	192,452	192,452
Total Expenses	633,332	518,704	2.82%	495,815	518,704	-4.41%	516,704	518,704

Business Insights:

Q2 revenue targets were off by 11.08%

Marketing spending was under budget at 9.13%

1.3. Descriptive Analytics

Process Details

Performance Analysis

What happened?

Associated with data visualization via reports, dashboards, and scorecards that facilitate decision making.

- State business metrics
- Identify data required
- Extract and prepare data
- Analyze data
- Present data

1	PLANNING ANALYTICS	What is your plan?
2	DESCRIPTIVE ANALYTICS	What happened?
3	DIAGNOSTIC ANALYTICS	Why did it happen?
4	PREDICTIVE ANALYTICS	What happens next?
5	PRESCRIPTIVE ANALYTICS	What should you do next?

Case Study: Chicago Bikeshare

Performance Analysis

What happened?

The Q2 revenue is down.

What do I need to do?

- Ask questions to understand the history
- Identify past issues
- Interact with data to ask new questions
- Find answers quickly

Performance Dashboard

ACTUAL $7.8 M

BUDGET $8.7 M

VARIANCE -10.6%

$1.6 MTD | $4.8 QTD $9.2 YTD

$1.6 MTD

-5.9% QTD -6.4% YTD

$7.8M USD revenue for the month is **10.6% below plan**

Underperformance for **4 months** out of 6

Revenue

Districts	Actual revenue	Budget revenue	Variance $	Variance %
ADACHI	$22.4 K	$12.13 K	$10.27 K	84.6%
AMADORA	$21.4 K	$24.97 K	-$3.54 K	-14.2%
BERLIN	$14.4 K	$12.45 K	$1.95 K	15.7%
BOUCHERVILLE	$28.1 K	$32.72 K	-$4.6 K	-14.1%
BRISBANE	$21.2 K	$12.84 K	$8.36 K	65.1%
CASTLE HILL	$26.9 K	$25.65 K	$1.23 K	4.8%
CLAREMONT	$13.7 K	$12.6 K	$1.08 K	8.6%
DEMPSEY	$74.5 K	$57.89 K	$16.62 K	28.7%
DIASPORA	$38.9 K	$37.78 K	$1.12 K	3.0%
DOWNTOWN	$1,361.9 K	$1,732.93 K	-$371.07 K	-21.4%

Expense

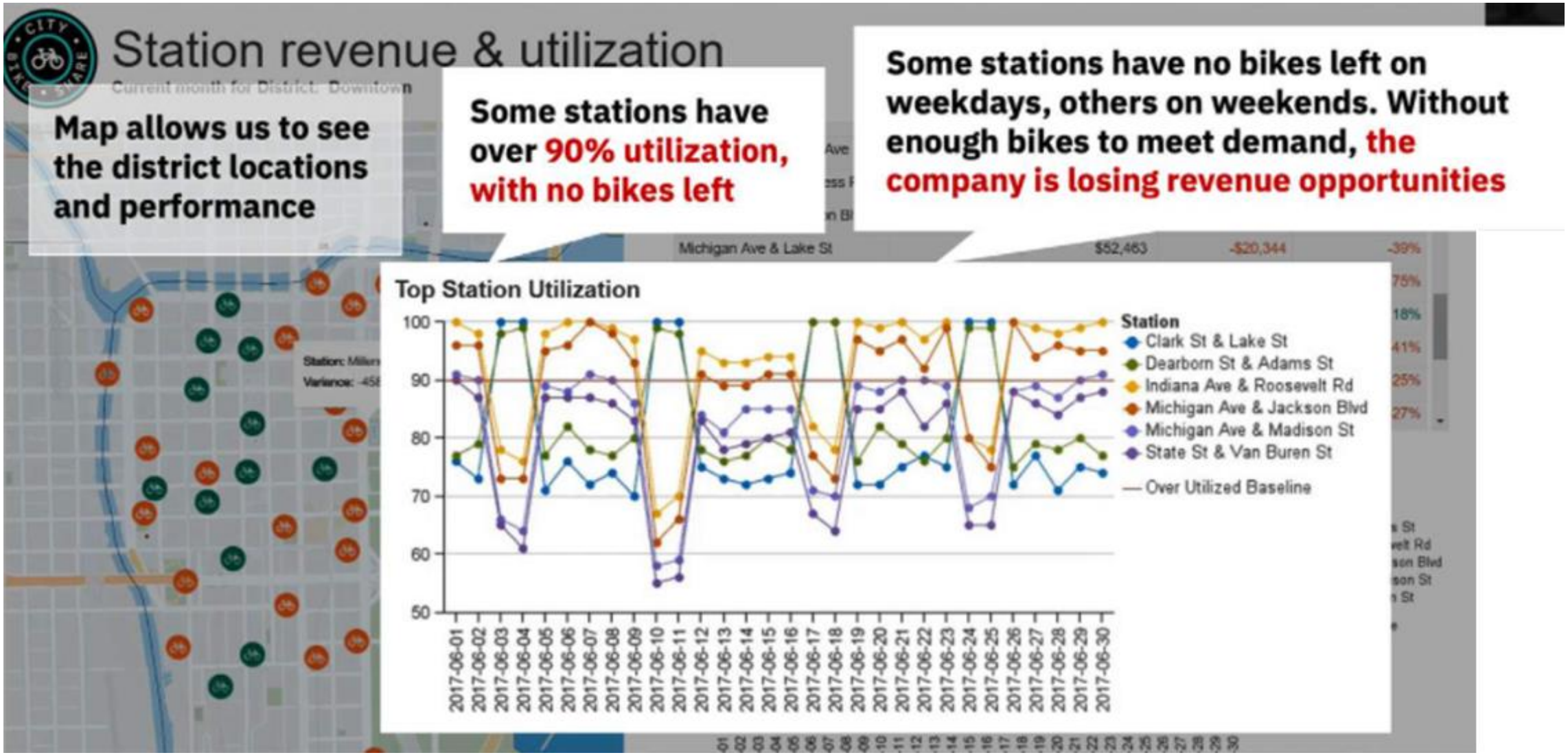
Districts	Actual expense	Budget expen		
ADACHI	$17.48 K	$9.49 K	$7.99 K	84.1%
AMADORA	$10.48 K	$12.23 K	-$1.76 K	-14.4%
BERLIN	$14.73 K	$12.78 K	$1.95 K	15.3%
BOUCHERVILLE	$27.52 K	$32.13 K	-$4.62 K	-14.4%
		$12.96 K	$8.36 K	64.5%
		$21.73 K	$0.97 K	4.5%
		$8.22 K	$0.66 K	8.3%
		$24.65 K	$7.01 K	28.5%
		$16.61 K	$0.46 K	2.8%
DOWNTOWN	$123.99 K	$172.9 K	-$48.95 K	-28.3%

Districts highlighted in blue are underperforming

Descriptive Analytics

What happened? Performance Understanding

Chicago Bikeshare



1.4. Diagnostics Analytics

Process Details

Identify Business Causes

Why did it happen?

Uncover details such as the frequency of events, the cost of operations and the root cause of failures

- Identify anomalies
- Drill into the analytics (discovery)
- Determine causal relationships

1	PLANNING ANALYTICS	What is your plan?
2	DESCRIPTIVE ANALYTICS	What happened?
3	DIAGNOSTIC ANALYTICS	Why did it happen?
4	PREDICTIVE ANALYTICS	What happens next?
5	PRESCRIPTIVE ANALYTICS	What should you do next?

Case Study: Chicago Bikeshare

Identify Business Causes

Why did it happen?

Identified an issue with bike supply Downtown. Discovered that the issue is causing revenue loss.

You need to:

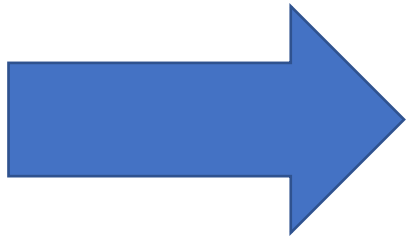
- Get to the bottom of your revenue problem and understand real drivers.
- Set aside your assumptions to find an accurate solution quickly.
- Assess what is driving different levels of demand at different bike stations using all your relevant data source.

Diagnostic Analytics

Why did it happen? Identify business causes

Chicago Bikeshare

See all your available data sources,
Including weather information,
To help understand what is driving demand.



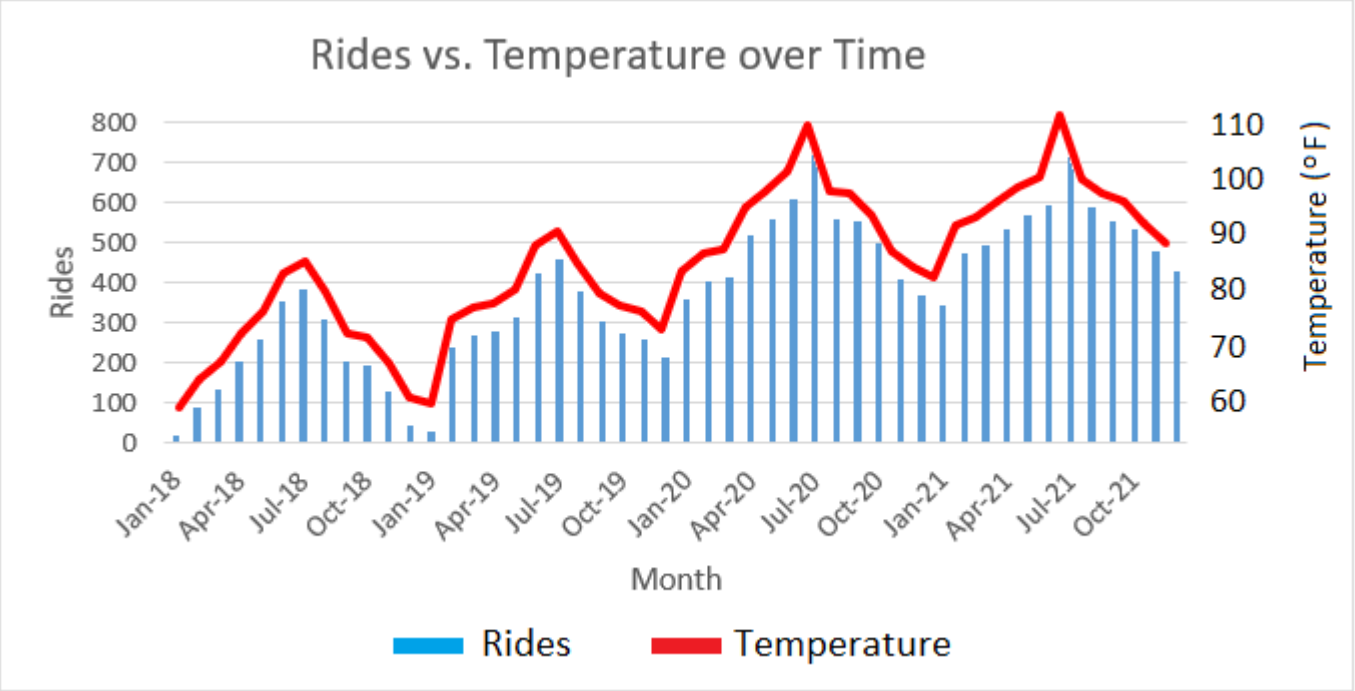
Noticed that the top drivers for the number of rides are temperature range and month, as well as humidity and dew point.

Diagnostic Analytics

Why did it happen? Identify business causes

Chicago Bikeshare

Found:
temperature correlation with the bikes ride demand



1.5. Predictive Analytics

Process Details

Predict Using Data Models

What happens next?

Using descriptive data accumulated over time, predictive analytics utilizes models for predicting events.

- Identify business outcomes
- Determine data required to train the prediction model
- Determine types of analysis
- Validate results
- Test predictions on performance

1	PLANNING ANALYTICS	What is your plan?
2	DESCRIPTIVE ANALYTICS	What happened?
3	DIAGNOSTIC ANALYTICS	Why did it happen?
4	PREDICTIVE ANALYTICS	What happens next?
5	PRESCRIPTIVE ANALYTICS	What should you do next?

1.6. Prescriptive Analytics

Process Details

Solution Optimization

1	PLANNING ANALYTICS	What is your plan?
2	DESCRIPTIVE ANALYTICS	What happened?
3	DIAGNOSTIC ANALYTICS	Why did it happen?
4	PREDICTIVE ANALYTICS	What happens next?
5	PRESCRIPTIVE ANALYTICS	What should you do next?

What should you do?

Use advanced capabilities such as optimization and mathematical models to reveal not only recommended actions but also why they are recommended, along with any implications the actions might have.

- Input data from model prediction
- Combine data models with rules
- Provide constraint-based optimization
- Implement for better decisions

Case Study: Chicago Bikeshare

Solution Optimization

What should you do?

- Identified an issue with bike supply Downtown.
- Discovered that the issue is causing REVENUE LOSS.
- Used all relevant data sources to assess that weather is driving levels of demand at different stations.
- Ran predictive analytics to identify future demand.

You need to:

- Enact an improved plan that optimizes bike allocation on a day-to-day, location-by-location basis.

Predictive Analytics

What should you do? -Solution Optimization

- An optimization system recommends the number of bikes needed to satisfy the predicted demand.
- In this case, Downtown will require an extra number of N bikes.

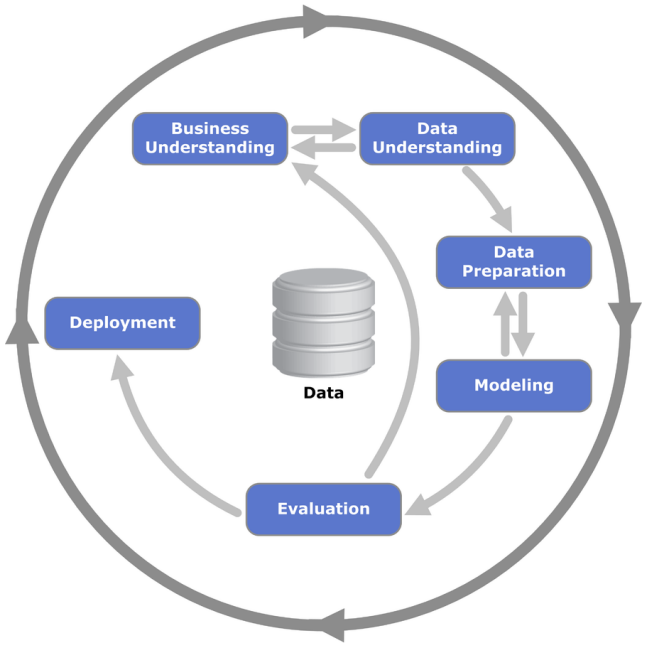
2. Data Analytics Methodologies

These methodologies provide the data scientist with a framework for how to proceed with whatever methods, processes and heuristics will be used to obtain answers or results.

Existing Methodologies

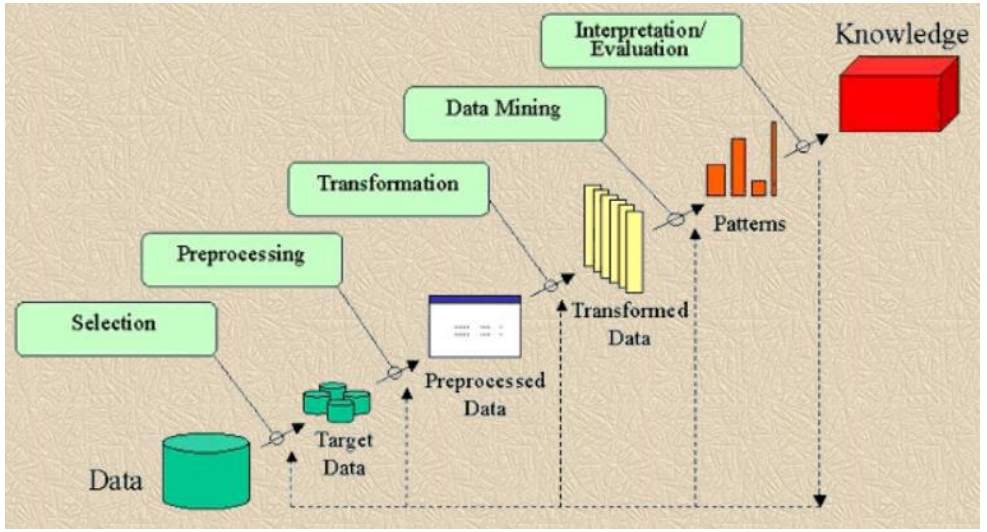
CRISP-DM

Cross-industry standard process for data mining



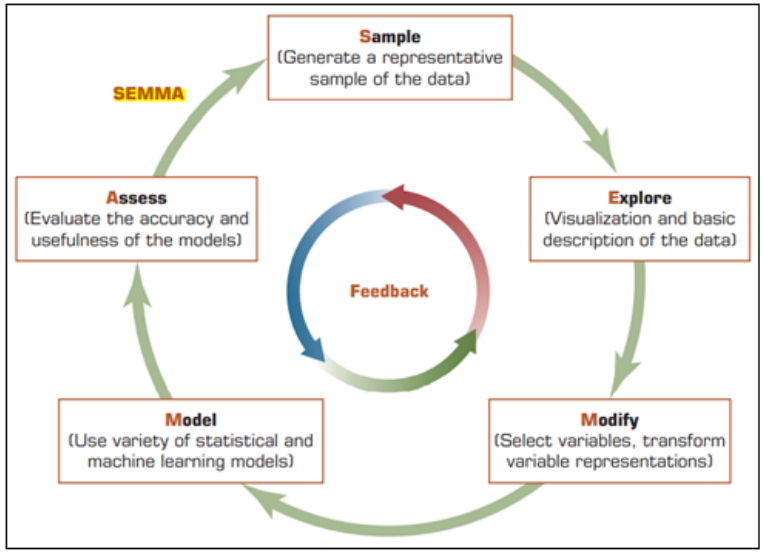
KDD

Knowledge Discovery in Databases



SEMMA

Sample, Explore, Modify, Model, and Assess



Data Analytics Methodologies

Which one to use?

Ranking:

1. CRISP-DM
2. SEMMA
3. KDD

Common characteristics:

- Data gathering
- Data transformation
- Data modeling
- Model evaluation

Differences:

- CRISP-DM includes business understanding

The Need for a Data Science Methodology

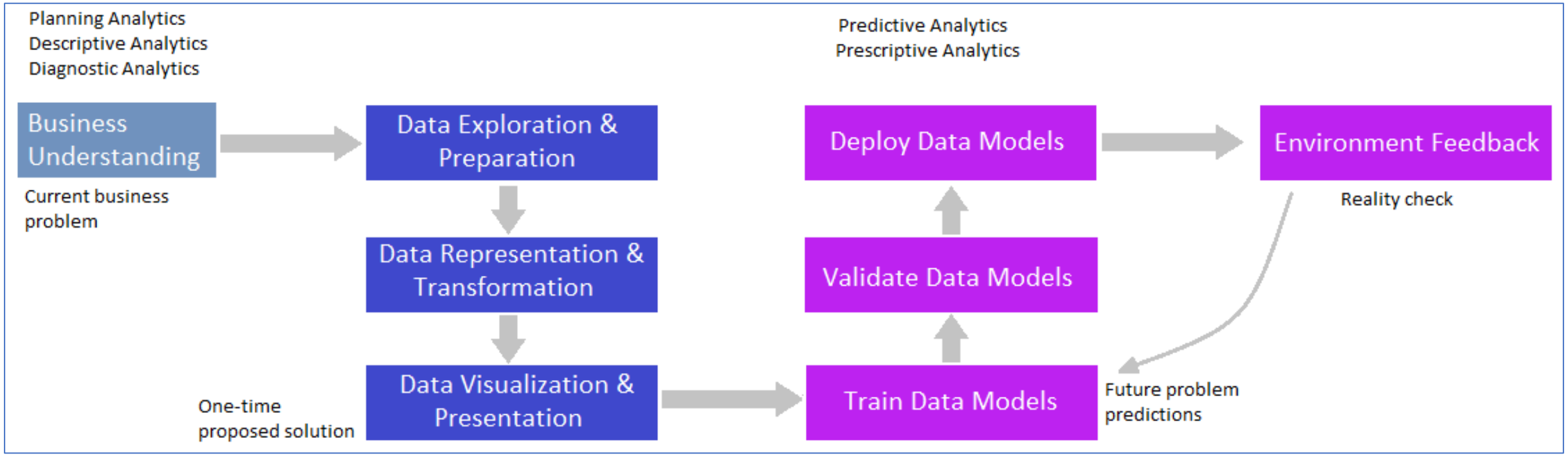
Data Scientists need a foundational methodology hat also addresses **unstructured data** and prescriptive analytics.

The methodology needs to emphasize new practices in Data Sciences such as:

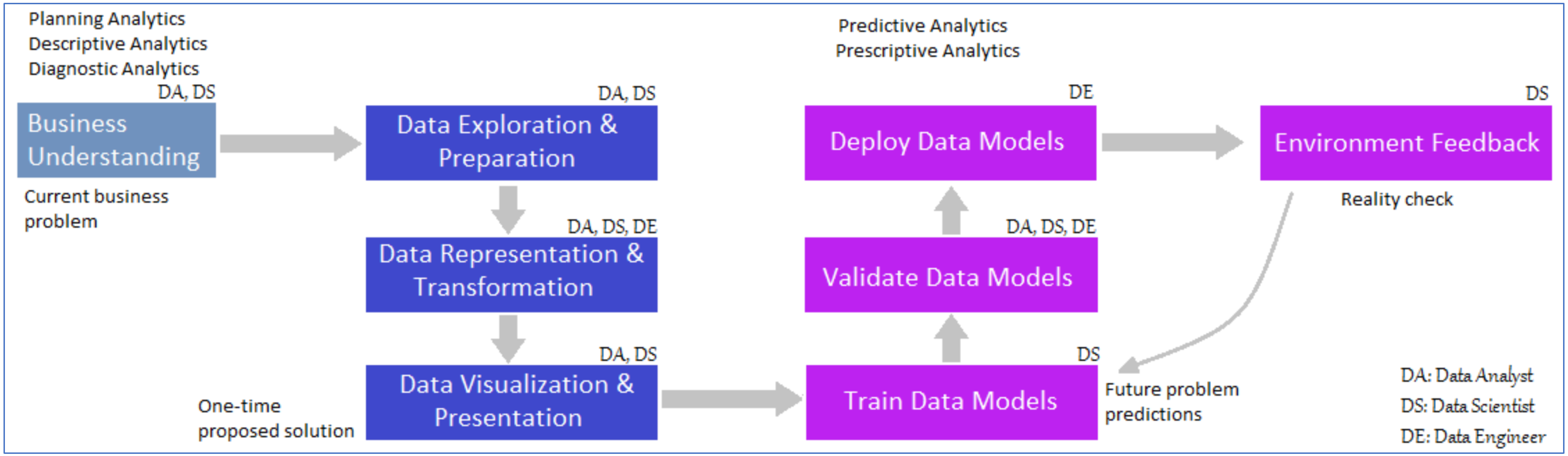
- Use of very large volumes of data
- Incorporation of Machine Learning capabilities into prescriptive modeling and automation

3. Data Science Method

3.1. The Phases of the Data Science Method



Data Analytics Methodology



3.2. The Method Applied to a Practical Scenario

Project Objective

The Fraud Claims Dept. Manager (Business sponsor)
Is trying to predict fraud or suspicious activity and their pattern to help drastically reduce losses due to frauds.

Business goals:

He asks the team to analyze the insurance claims in search for fraudulent behaviors.

Tasks:

- Identify auto insurance claims filed after the expiration of the insurance policy.
- Claims filed after the license expiration date.
- Excessive claim amount, which is over $10,000 in value.

Data Requirements

The Data Analyst will be involved.

Business Needs:

- Identify auto insurance claims filed after the expiration of the insurance policy.
- Identify claims filed after the license expiration date.
- Identify excessive claim amount , which is over $10,000 value.

Requirements:

- Datasets must include details about the insurance policy.
- Dataset should include driver license details.
- Dataset must give access to a large sample of actual insurance claims.

Data Collection

After a week going through a back and forth with the Database admin to confirm the size and scope of the sample dataset and getting approvals to handle the personal sensitive information contained in the file, the Data Analyst finally got access to the sample dataset.

Privacy Considerations:

- The dataset contains sensitive personal data, subject to data privacy regulations and will need to be removed.

Data Cleansing

Recap: Now that the sensitive data has been removed, it is time to process the information for analysis.

Next: The data types of the columns of the data file will need to be checked. Some types will need to be converted (for example, dates represented as strings will need to be converted to date format)

Data Understanding

During the Data Understanding phase, the business sponsor, the data analysts and the data scientist need to go over all the columns of the file and establish their meaning, data format, and valid range of values.

Some variables might be coded:

INCIDENT_CAUSE:
1 = Driver error
2 = Natural causes
3 = Other driver error
4 = Crime
5 = Other causes

CLAIM_STATUS:
1 = Open
2 = Approved
3 = Paid
4 = Flagged for fraud
5 = Denied
6 = Appeal

POLICE_REPORT:
1 = There was police report
0 = No police report

Algorithm Alignment

During the phase of algorithm alignment, we need to identify the more suitable algorithms to analyze the potential 3 fraud scenarios described.

Some of the methods and tools are:

- Logistical regression
- Scatter plots
- Other graphs
- Statistical analysis
- Create columns with Boolean values to be used as flags

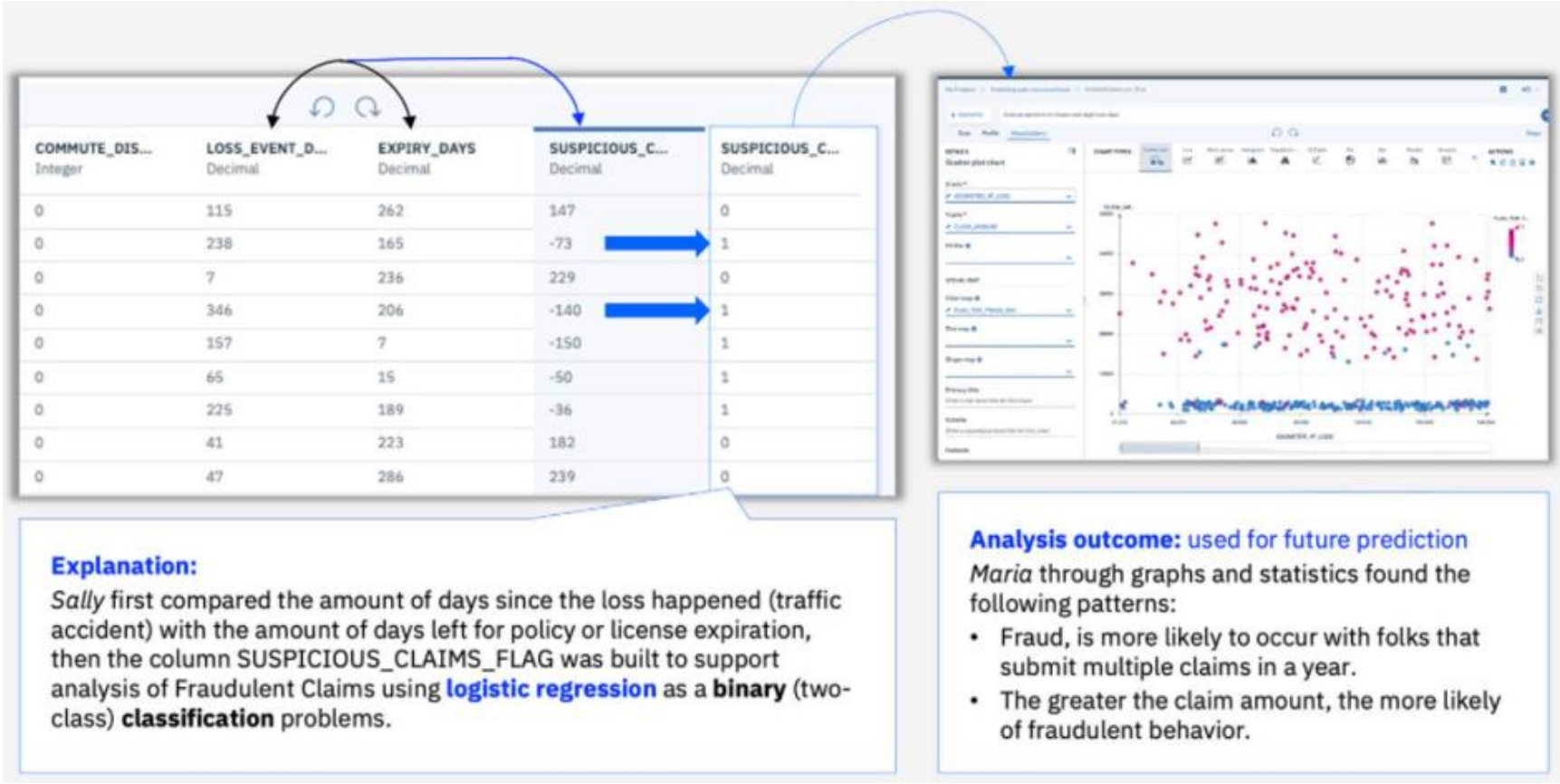
Data Formatting

During the phase of algorithm alignment, the Data Analyst uses tools applying mathematics to aggregate the information, facilitating the application of statistical analysis to the sample data.

Usually, additional temporal columns must be added to the dataset to serve as intermediate steps before a final prediction is made.

Statistical Analysis and Exploratory Data Visualization

During the phase of statistical analysis and exploratory data visualization, the descriptive statistics of each data column are calculated.
Furthermore, different graphs are utilized to visualize each data column, such that they can be better interpreted.

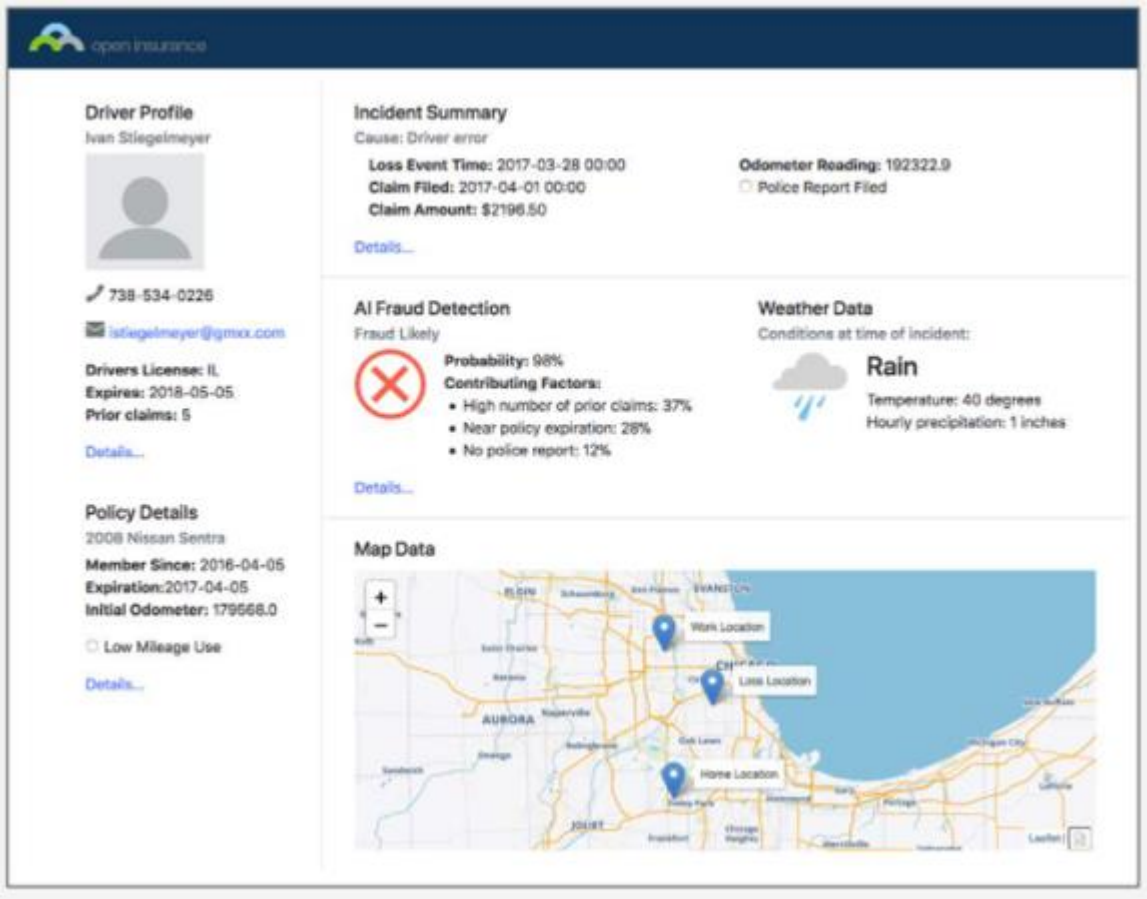


Decision Support System

Creating a dashboard can reduce the time to validate fraudulent claims.
A Fraud Investigation Dashboard will represent a life chart that is refreshed with current data automatically.

Teams involved:

- IT department
- Data Science team



4. Integrated Environments for Data Science Projects

4.1. Industry Applications for Cloud-based Data Management

Why are enterprises struggling to capture the value of data?

Data	Governance	Skills	Tools & Infrastructure
• Data resides in silos & difficult to access	• If the data isn't secure, self-service isn't a reality	• Data Science skills are in low supply and high demand	• Need an environment that enables a "fail fast" approach
• Unstructured and external data wasn't considered	• Challenge understanding data lineage and getting to a system of truth	• Nurturing new data professionals is challenging	• Discrete tools present barriers to productivity

What is the solution?

An integrated development environment for Data Science projects

Closing the Gap for Data Science Adoption

Problem:

Gap between data experts and domain experts

Solution:

Data Science Cloud platforms close the gap with a unified experience to create new insights from knowledge contained in the data, enabling multidisciplinary teams across the organization to collaborate.

4.2. Roles-integrated Environment

Data Scientist

Her Job:

Transform data into knowledge to solve business problems

What she does:

- Follows the project end-to-end
- Discover content from multiple data sources
- Use popular statistical libraries
- Runs experiments to build custom models that solve business problems.
- Use techniques such as Machine Learning or Deep Learning and works with Sally to validate success of trained models.

Data Analyst

Her Job:

Captures domain knowledge for successful business alignment

What she does:

- She works with the domain experts to understand the business problems
- Provides domain knowledge that Maria and Tom uses to develop a custom data models
- Sally uses data refinery tools to cleanse and curate the raw data
- Makes data visualizations to depict insights to sponsor users

Data Engineer

His Job:

Architects how data is organized and ensures operability

What he does:

- Uses tools for data munging and data wrangling to develop pipelines that Maria can use to apply statistical methods and build models.
- Works with Maria to transform research models into production quality systems.
- Builds data infrastructure and ETL pipelines. Works with Spark, Hadoop, and HDFS.

AI Developer

His Job:

Builds AI application that meet the requirements of the business.

What he does:

- Starts PoCs which includes gathering content, dialog building and model training
- Focus is on app building for the team or company to use. Will handle ML Ops as needed
- Front-end, back-end, full stack, mobile or low-code developer

Roles integrated environment

SciKit Learn

Tensor Flow

[7]
You need framework for **Machine Learning** and likely framework for **Deep Learning**
Scikit, TensorFlow, Keras, Torc, Theano, etc.

---

[6]
Choose your scientific computing and statistic packages:
SciPy, NumPy are widely utilized

NumPy

SciPy

---

[5]
Choose your visualization and plotting tools:
Matplotlib, PixieDust are top market trends

Pixie Dust

Open CV

MatPlotLib

---

[4]
Choose your data munging libraries and tools:
Pandas is a very flexible library under the python framework

Pandas

---

[3]
Choose your programming language:
python is most versatile, R and **Scala** are mostly specialized statistical packages

Packages in R

Python

---

[2]
Now you need a development environment.
- Get **Jupyter Notebooks** (julia+python+R)
- Get it from **Anaconda** (www.continiuum.io)

Anaconda

Jupyter

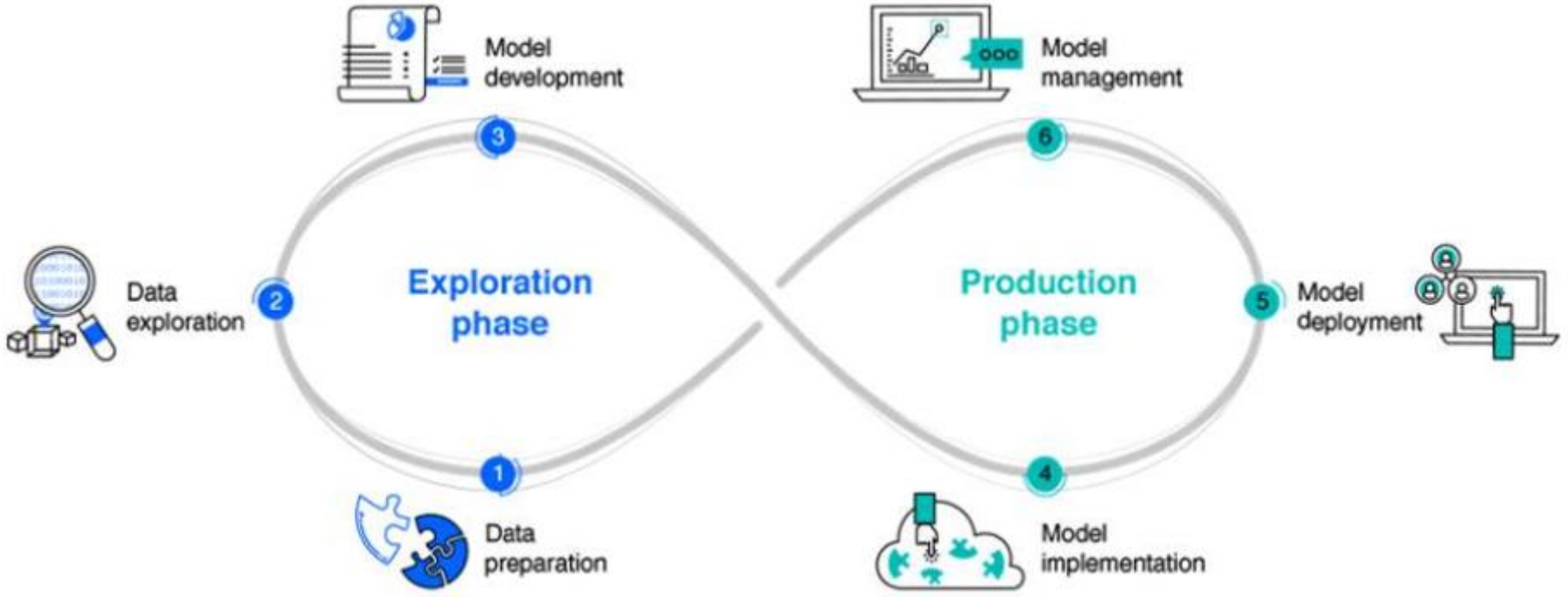
---

[1]
So you have your collected data:
Is it structured, semi-structured, unstructured, mix?

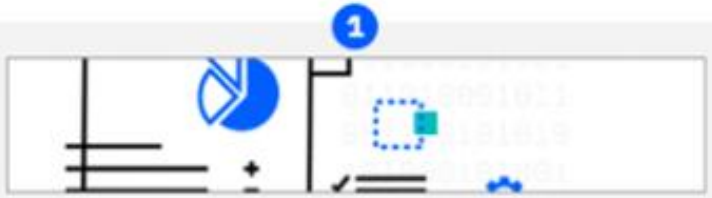
Text editor	MS Excel	DB2, MS SQL	Hadoop	MongoDB	Watson Studio	Sentiment
CSV	xls	RDBMS	DFS	JSON	images	text

5. Cloud-based Data Science Lifecycle

Bringing the data science methodology , roles and tools into a data science lifecycle.

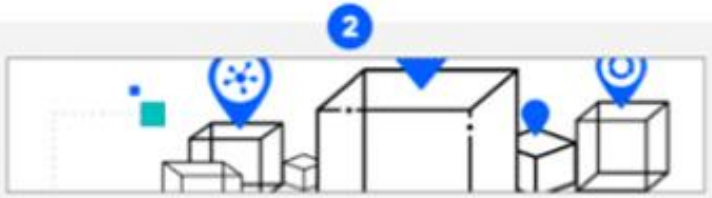


Exploration Phase



Data preparation

Data science projects start by asking the right business questions and collecting and preparing data. Success in the later phases is dependent on this early stage.



Data exploration

Once your data is in the right format to work with, you can conduct the next phase in the data analysis process. This initial exploration of the dataset is critical.

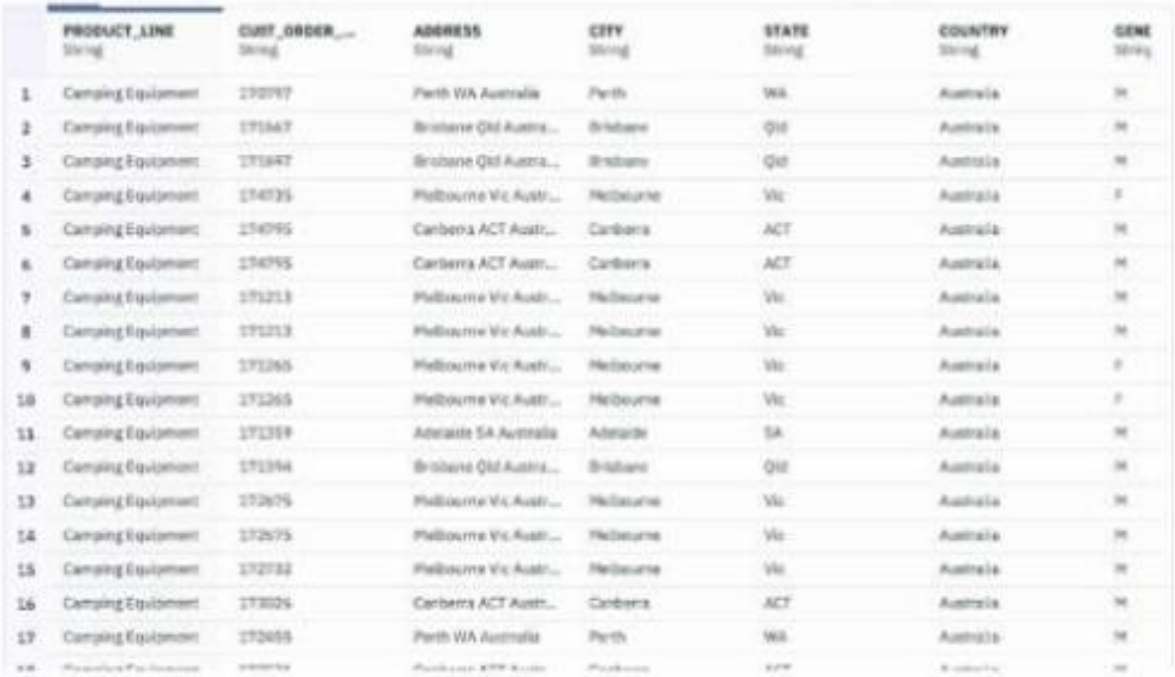
Model development

Here, the data you've prepared is brought into the data science toolset and the results begin to shed some light on previously identified business problems.

Exploration Phase: (1) Data Preparation

Uncover hidden insights in your data with data cleansing and refinery tools which will clean and transform the data for better processing.

Gain access to tabular view of your data, including visualizations and summary statistics that can help you uncover hidden insights by asking the right business questions of your data.

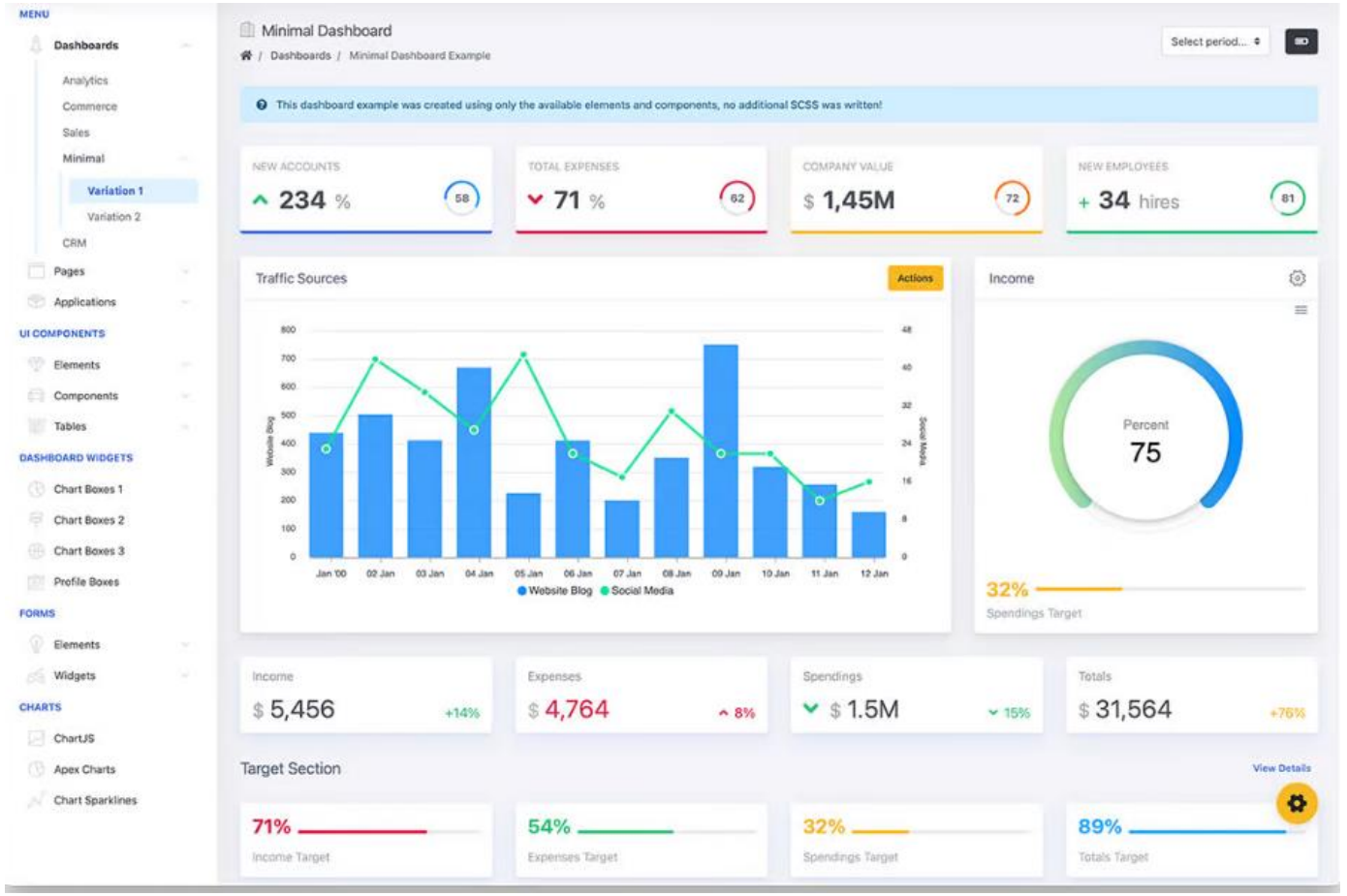


	PRODUCT_LINE String	CUST_ORDER_... String	ADDRESS String	CITY String	STATE String	COUNTRY String	GENE String
1	Camping Equipment	170797	Perth WA Australia	Perth	WA	Australia	M
2	Camping Equipment	171647	Brisbane Qld Austra...	Brisbane	Qld	Australia	M
3	Camping Equipment	171647	Brisbane Qld Austra...	Brisbane	Qld	Australia	M
4	Camping Equipment	174735	Melbourne Vic Austr...	Melbourne	Vic	Australia	F
5	Camping Equipment	174795	Canberra ACT Austr...	Canberra	ACT	Australia	M
6	Camping Equipment	174795	Canberra ACT Austr...	Canberra	ACT	Australia	M
7	Camping Equipment	171213	Melbourne Vic Austr...	Melbourne	Vic	Australia	M
8	Camping Equipment	171213	Melbourne Vic Austr...	Melbourne	Vic	Australia	M
9	Camping Equipment	171265	Melbourne Vic Austr...	Melbourne	Vic	Australia	F
10	Camping Equipment	171265	Melbourne Vic Austr...	Melbourne	Vic	Australia	F
11	Camping Equipment	171359	Adelaide SA Australia	Adelaide	SA	Australia	M
12	Camping Equipment	171394	Brisbane Qld Austra...	Brisbane	Qld	Australia	M
13	Camping Equipment	172675	Melbourne Vic Austr...	Melbourne	Vic	Australia	M
14	Camping Equipment	172675	Melbourne Vic Austr...	Melbourne	Vic	Australia	M
15	Camping Equipment	172732	Melbourne Vic Austr...	Melbourne	Vic	Australia	M
16	Camping Equipment	173026	Canberra ACT Austr...	Canberra	ACT	Australia	M
17	Camping Equipment	172655	Perth WA Australia	Perth	WA	Australia	M

Exploration Phase: (2) Data Exploration

Using dashboarding services, you can produce stunning visualizations directly from your data in real time.

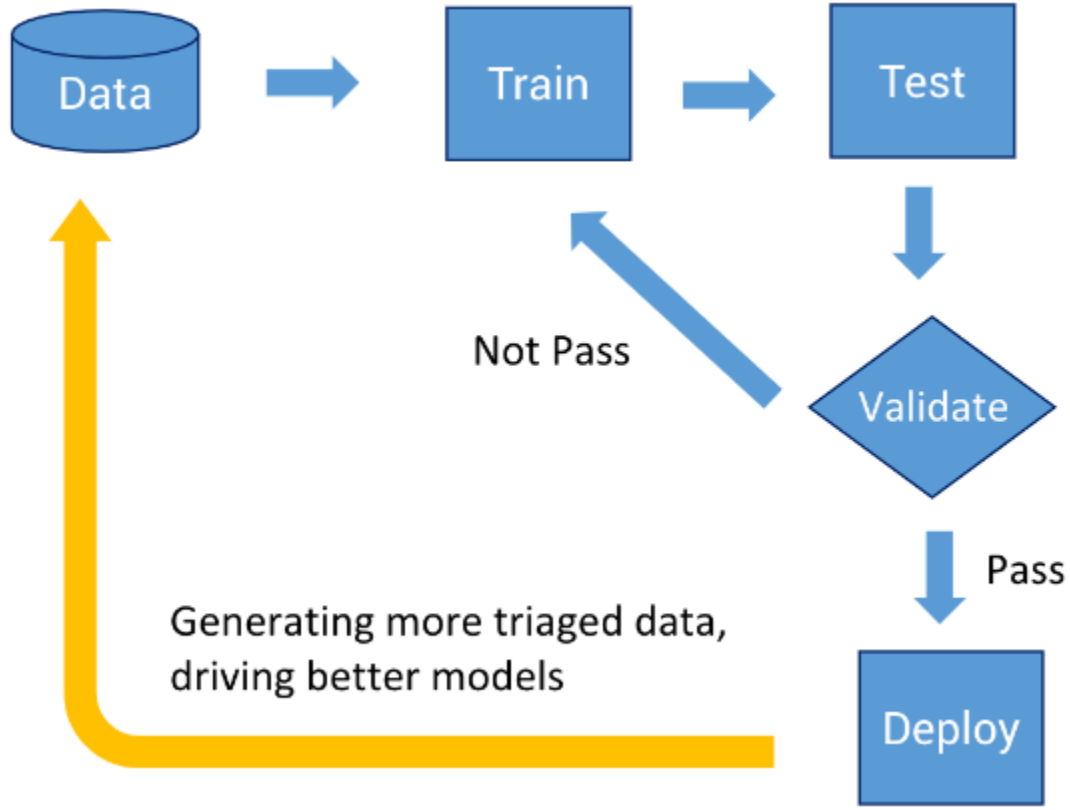
This allows you to illuminate previously unknown patterns, relationships or other actionable findings, and easily share them with your team.



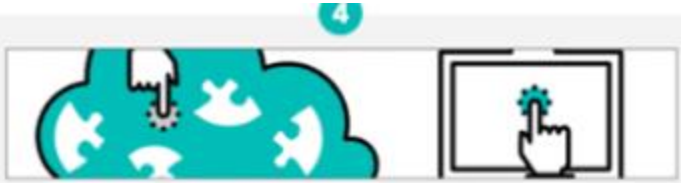
Exploration Phase: (3) Model Development

Test and deploy models, using customizable compute environments that scale up and down with your workflow.

Choose from various capacities of Anaconda Spark and GPU environments.



Production Phase



Model implementation

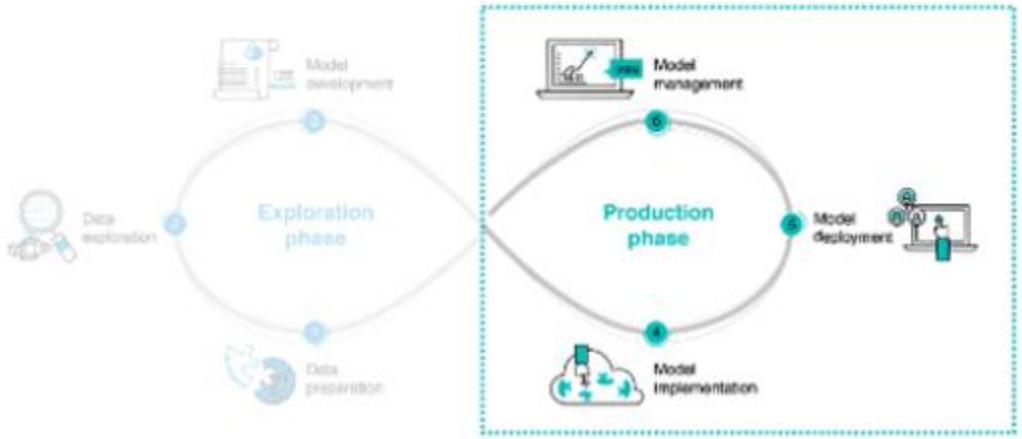
Once you've built and chosen your model, this stage helps you evaluate and understand its quality to ensure it fully addresses the business problem.

Model deployment

Upon development and approval by business sponsors, you're ready to deploy your model into the production environment or a comparable test environment.

Model management

Model management must be a continuous process to ensure optimal performance over time. This stage helps you monitor model creation, use and decay.



Production Phase

(4) Model Implementation:

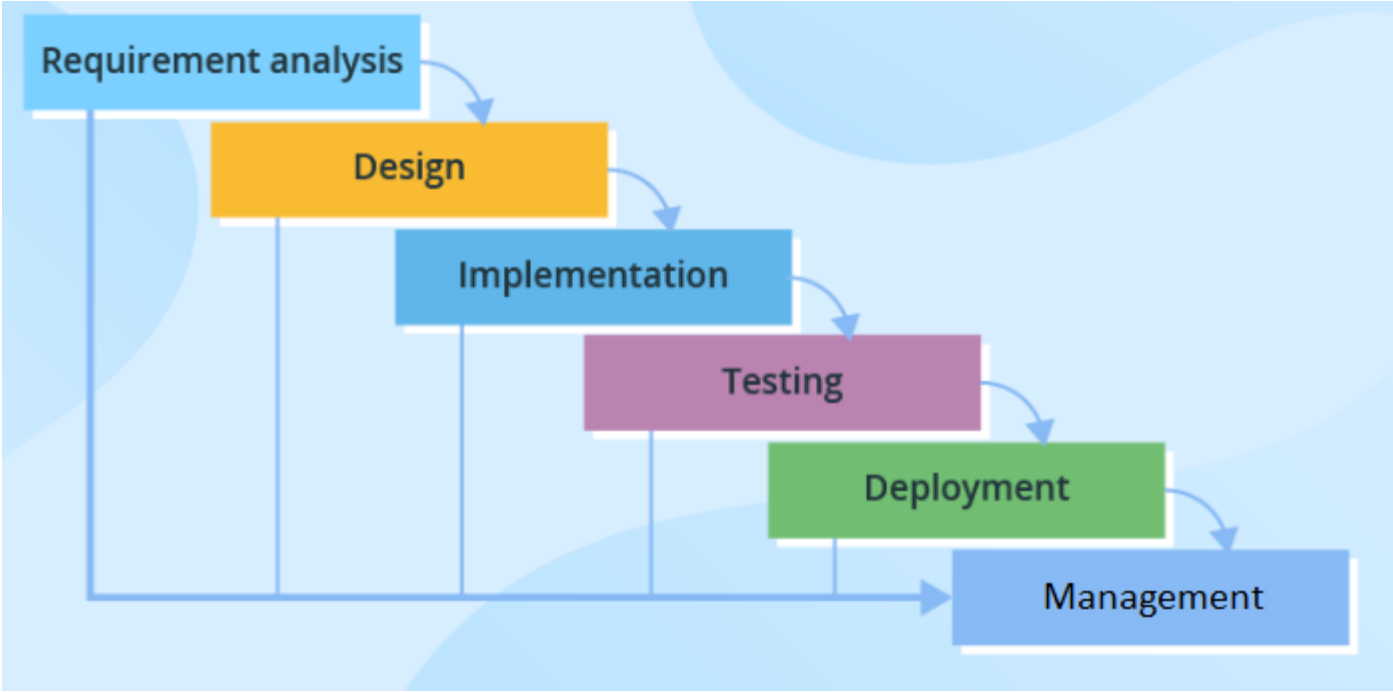
Improve your model's performance by visualizing fit between model and data.

(5) Model Deployment:

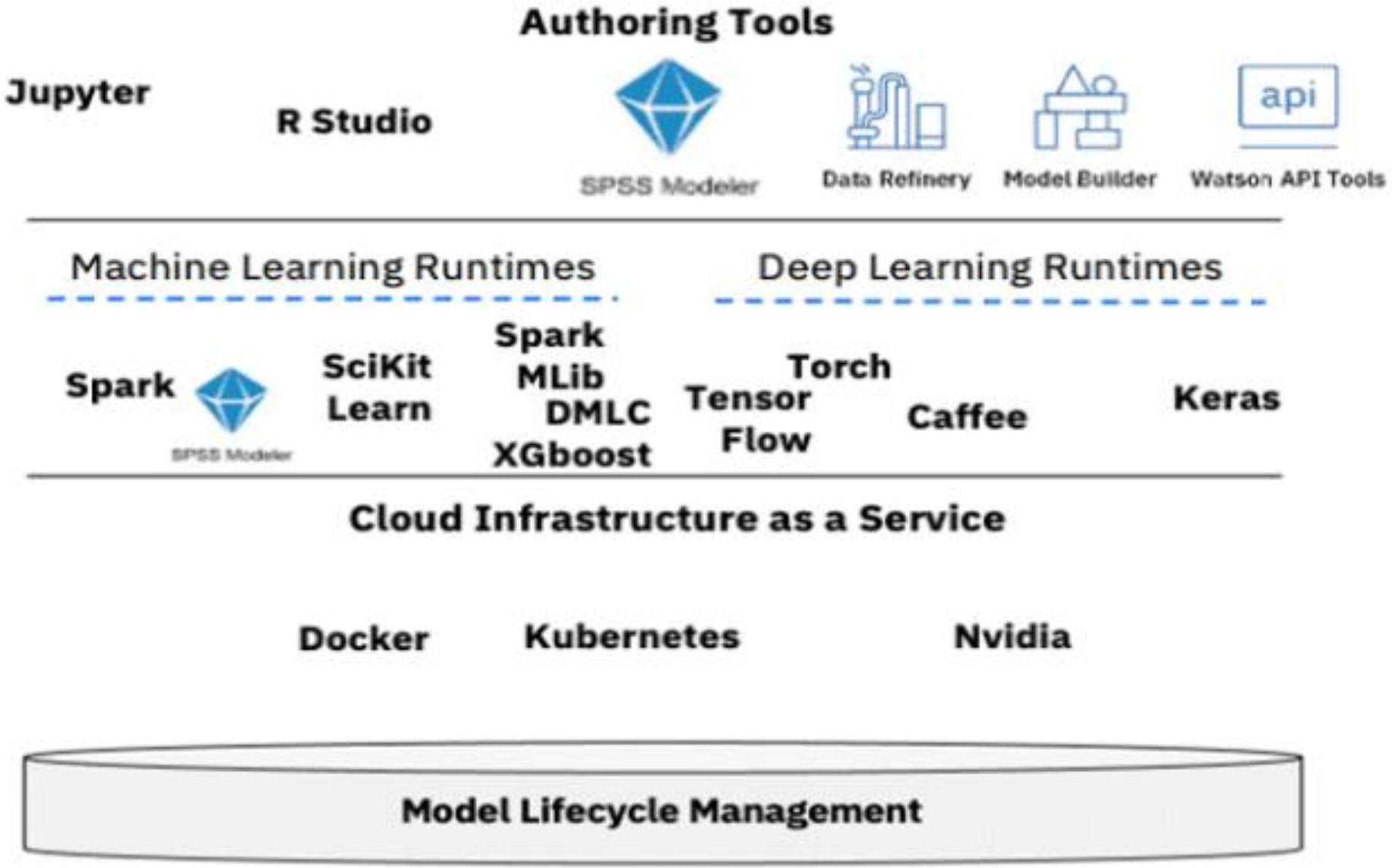
Once your model is ready, deploy and score it with the available services.

(6) Model Management:

Compare runs and conduct model hyperparameter optimization easily with deep learning experiments.



5. Data Science Capabilities on the Cloud



End of Lecture 3