

Summary of Data Mining and Analytics I

Overview of Data Mining

Types of Data Mining		
Basis for Comparison	Descriptive Mining	Predictive Mining
Basic	It identifies what happened in the past by analyzing stored data.	It describes what can happen in the future based on historic data.
Requirements	Data aggregation and data mining	Statistics and forecasting methods
Preciseness	Provides accurate data	Produce results that does not ensure accuracy.
Approach	Reactive	Proactive
Practical Analysis Methods	Standard reporting, query/drill down and ad-hoc reporting.	Predictive modelling, forecasting, simulation and alerts.
Examples	<i>Adverse events of a drug were explored by clustering the therapeutic classes; A data analyst receives detailed customer purchasing data and finds associations of any type among customers.</i>	<i>An automobile company scored customers for likelihood to return to buy a new model within the next 6 months; A credit card company offered a valued customer product for their card holder based on past card usage to determine the risk pattern; Road traffic was forecasted hourly</i>

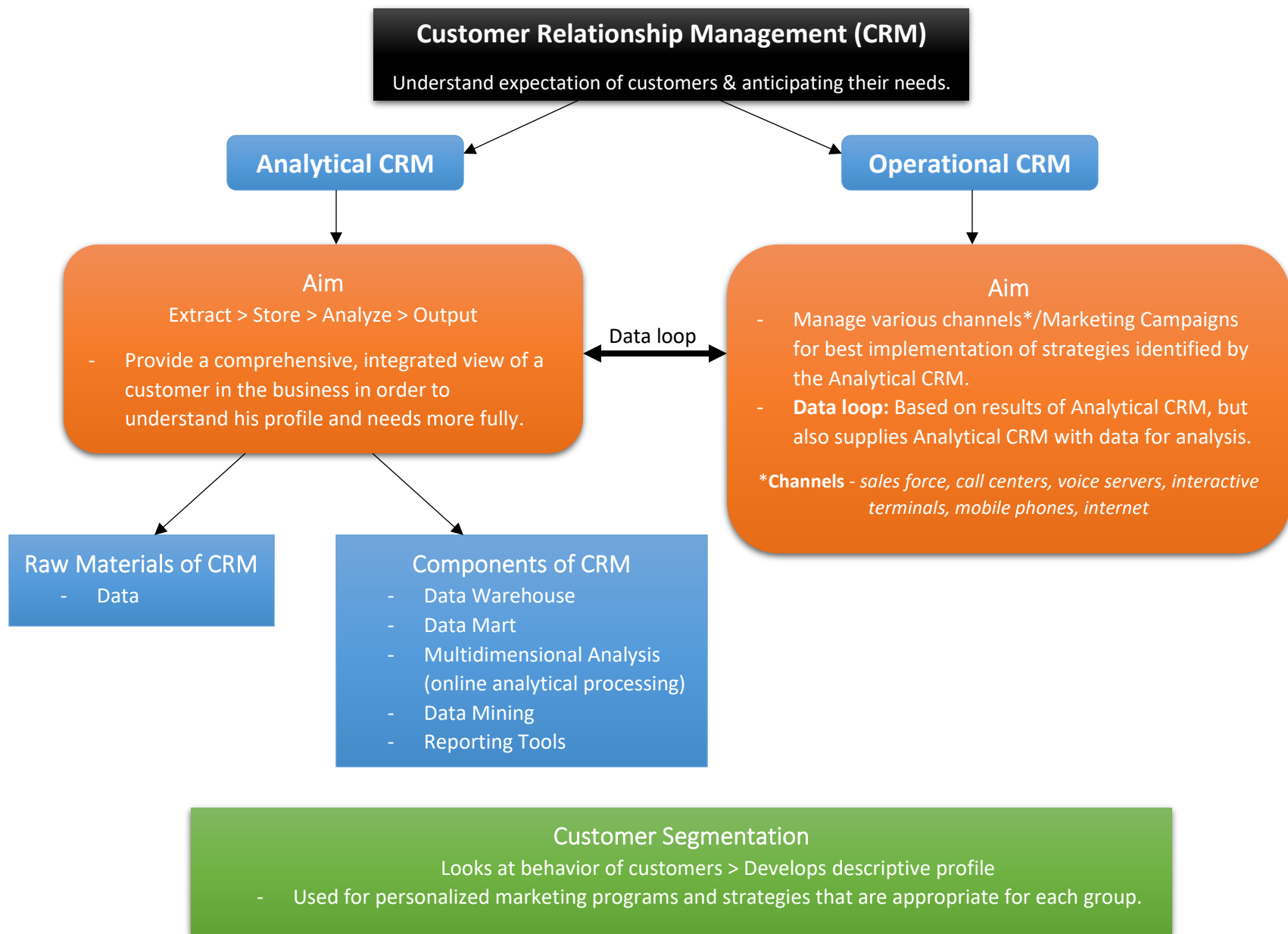
Data Mining Aims



Types of Values

Rare Value	Missing Value	Aberrant Value	Extreme Value
		Erroneous value an incorrect measurement, a calculation error, an input error or a false declaration. While extreme values are not always aberrant, aberrant values are not always extreme, and this makes them harder to detect, possibly requiring a thorough knowledge of the data.	

		However, it is essential to recognize them.	



To do:

- Qualitative, Explanatory data mining
- Types of data

Why might an online retailer mine the order history of its customers?

- To source new products

A data analyst is using analytical CRM to extract, store, analyze, and output relevant customer information. What is the first step within the analytical CRM phase that this analyst will be performing?

- Combining a customer's records to develop a holistic view.

Which feature of application development is unique to data mining?

The development phase cannot be completed in the absence of data.

