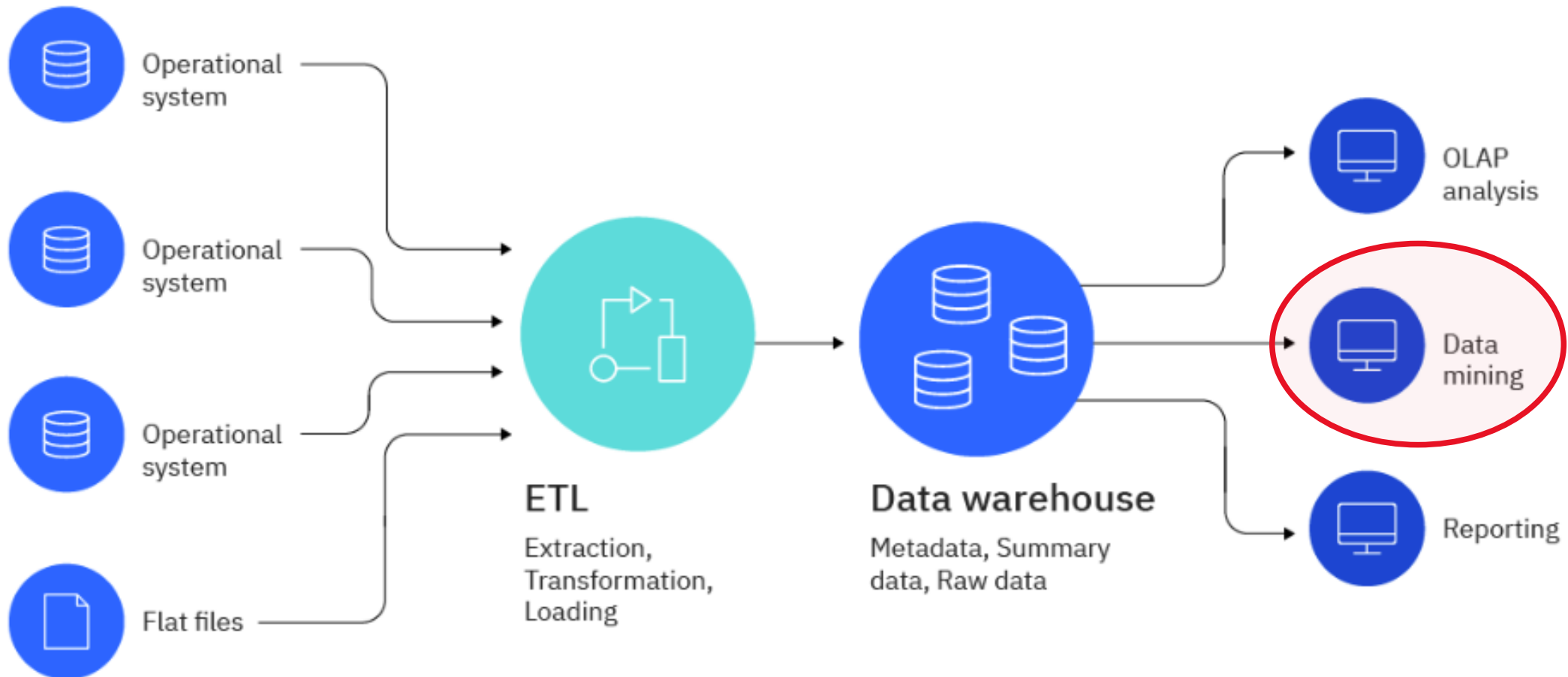


Data mining in Business Intelligence

Horvath Krisztina-Aliz

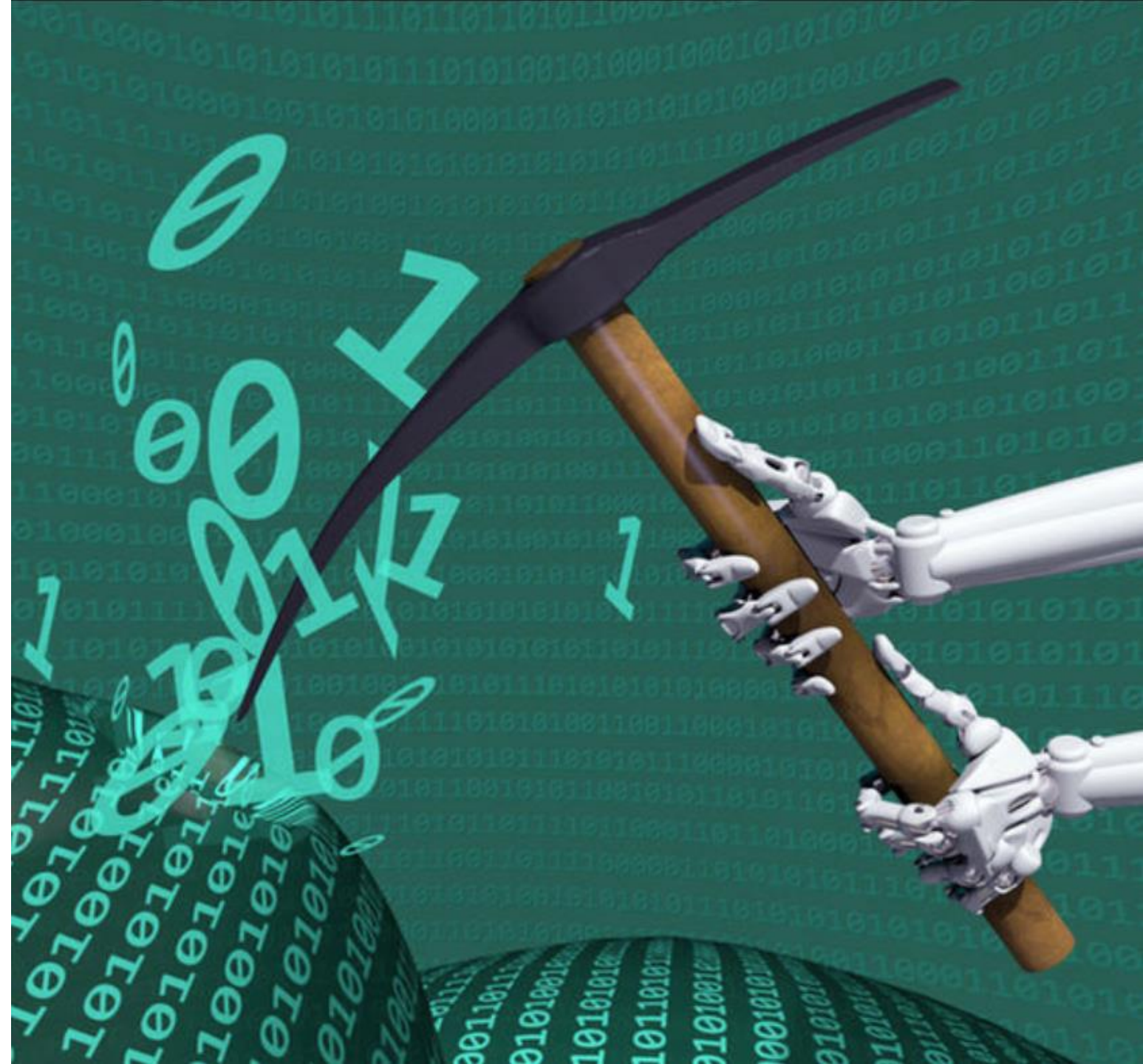


BUSINESS INTELLIGENCE PROCESS

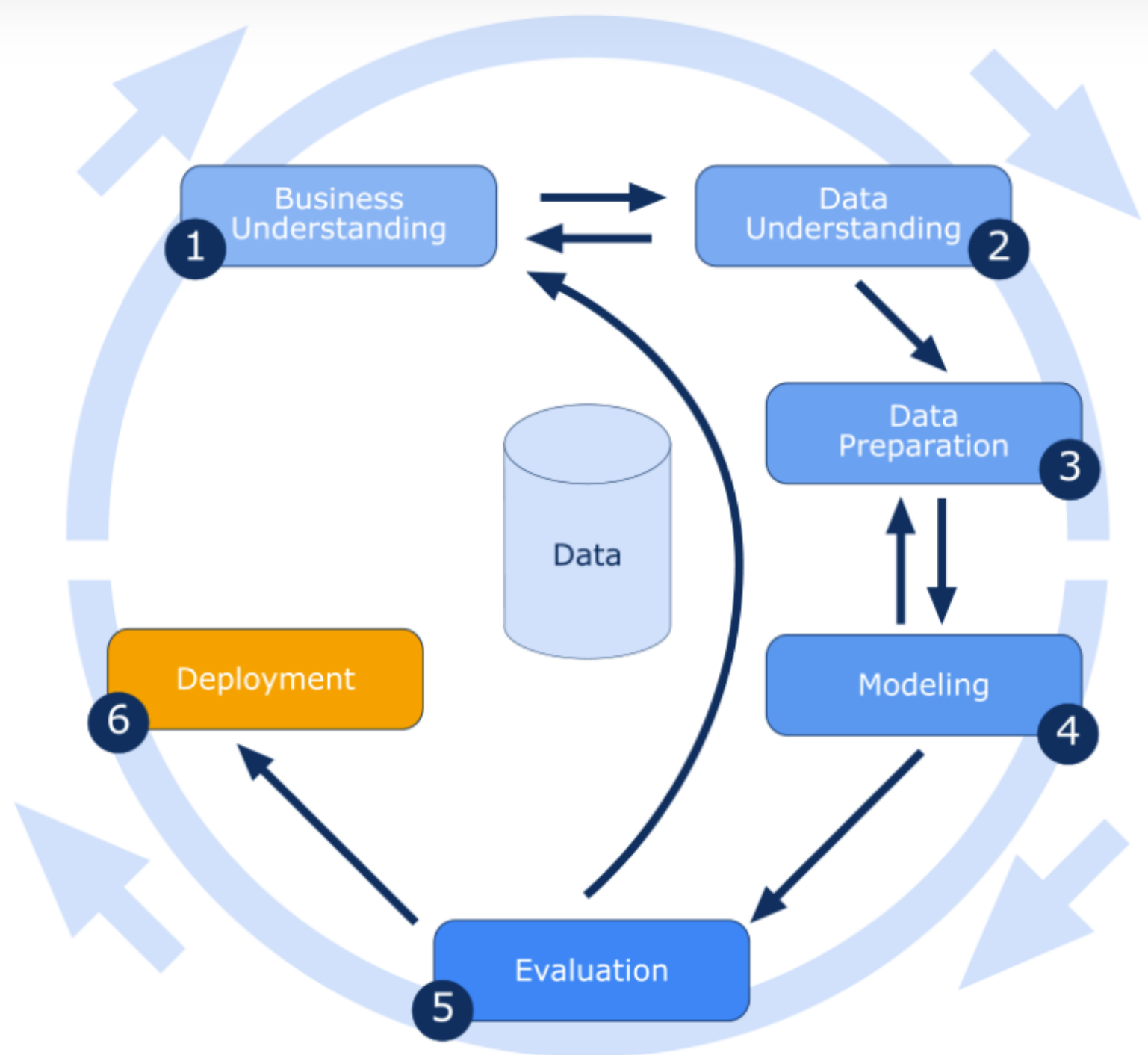


WHAT IS DATA MINING?

- Data mining, also known as knowledge discovery in data (KDD), is the process of uncovering patterns and other valuable information from large data sets.



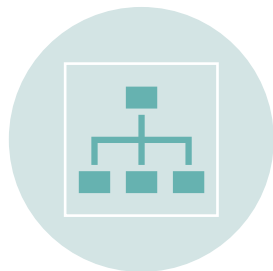
DATA MINING PROCESS



DATA MINING TECHNIQUES



REGRESSION



CLASSIFICATION



CLUSTERING

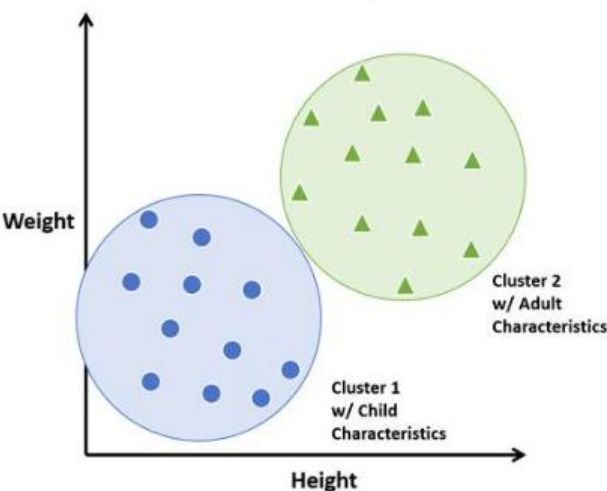
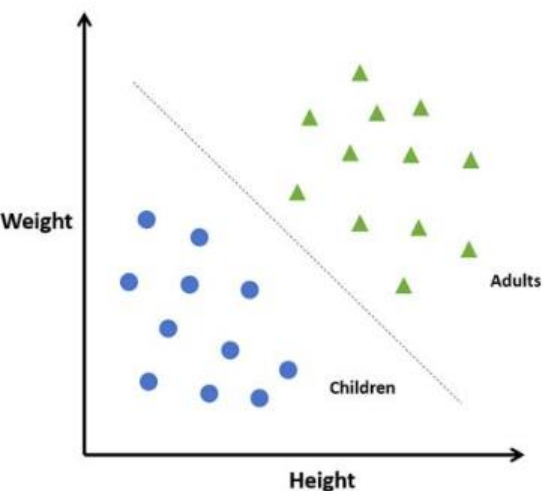


ANOMALY
DETECTION

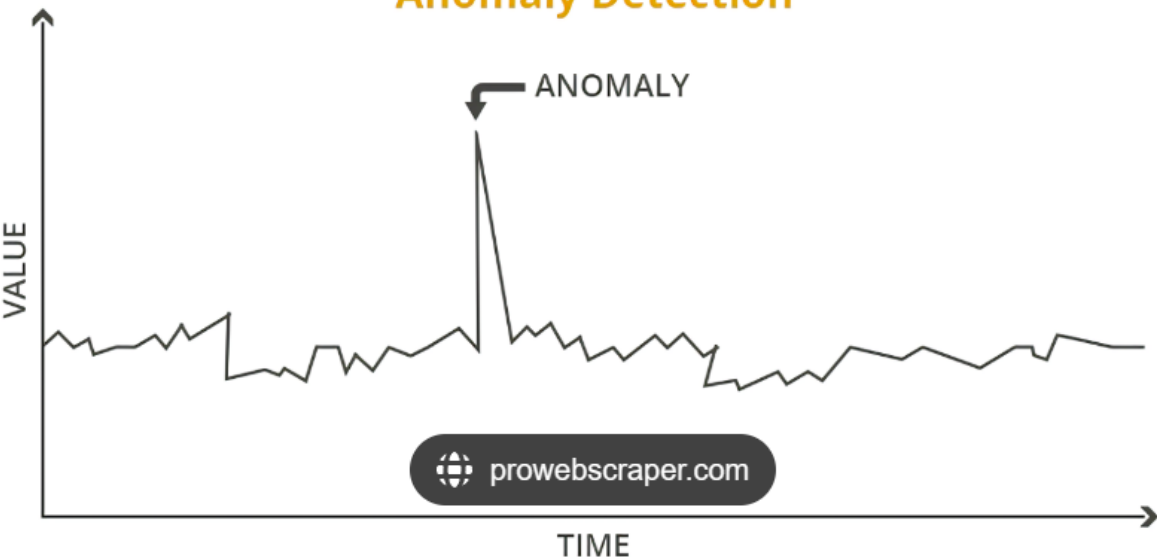
Classification

vs

Clustering



Anomaly Detection



DATA MINING QUERY IN SSIS

- Data Flow Task
- Connect to SSAS server where the Data Mining Model is deployed

Data mining query:

```
SELECT FLATTENED
  [Decision Trees Default].[Bike Buyer],
  (PredictProbability([Decision Trees Default].[Bike Buyer])) as [Probability]
From
  [Decision Trees Default]
PREDICTION JOIN
  @InputRowset AS t
ON
  [Decision Trees Default].[Marital Status] = t.[MaritalStatus] AND
  [Decision Trees Default].[Gender] = t.[Gender] AND
  [Decision Trees Default].[Yearly Income] = t.[YearlyIncome] AND
  [Decision Trees Default].[Total Children] = t.[TotalChildren] AND
  [Decision Trees Default].[Number Children At Home] = t.[NumberChildrenAtHome]
AND
  [Decision Trees Default].[House Owner Flag] = t.[HouseOwnerFlag] AND
  [Decision Trees Default].[Number Cars Owned] = t.[NumberCarsOwned]
```

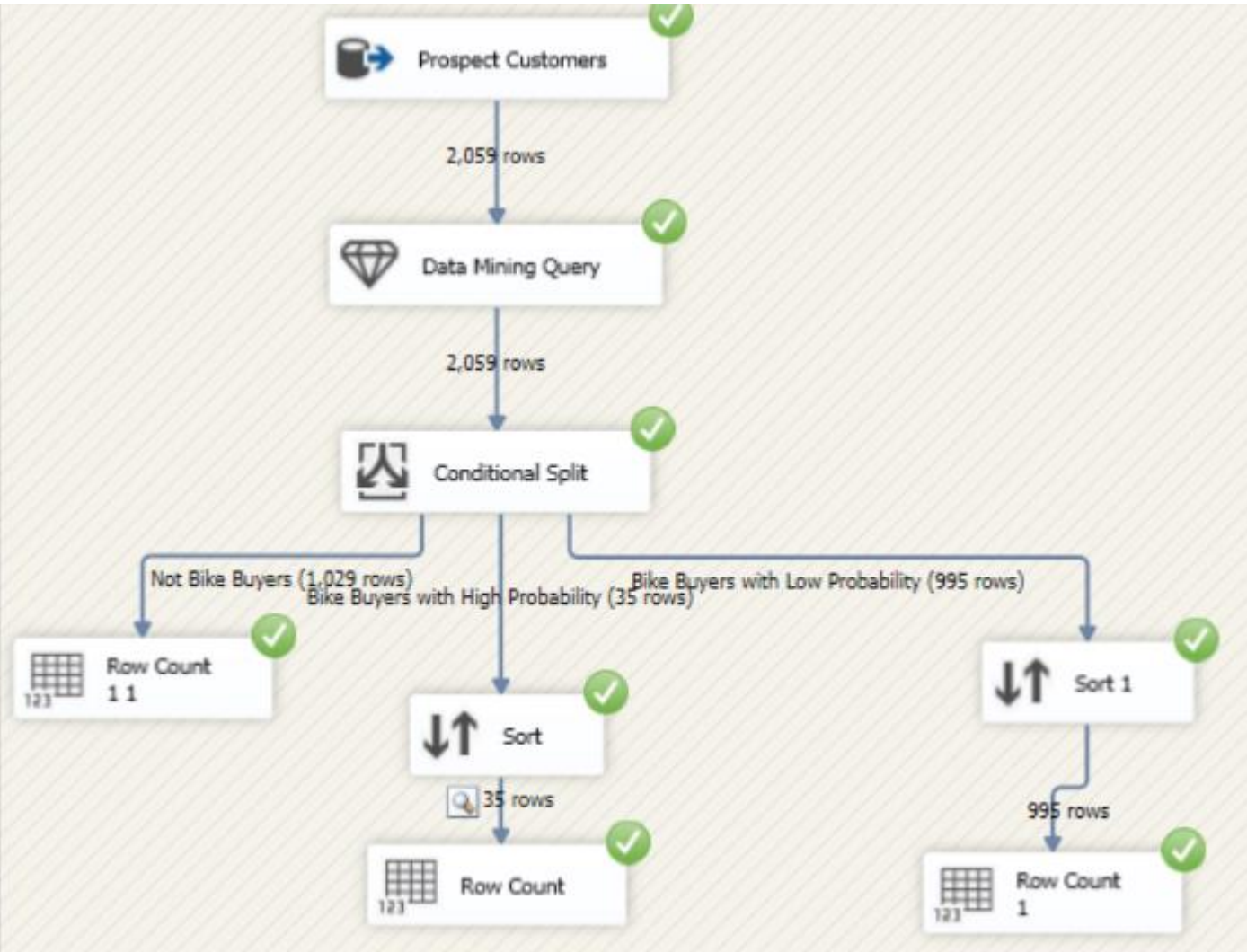

DATA MINING QUERY IN SSIS



Data Mining Query Data Viewer at Data Flow Task								
▶ Detach Copy Data								
FullName	Address	Phone	EmailAddress	Gen...	Yea...	Bike B...	Probability	
Mr. Adam Alexander	566 S. Main Cedar City U...	516-555-0187	aalexander@lucernepubli...	M	40...	0	0.691282668500688	
Mr. Adam Coleman	2964 St. George Dr. Col...	1 (11) 500 555-0...	acoleman@alpineskihouse...	M	20...	1	0.552484051807462	
Mr. Adam A Carter	4082 Roslyn Road Belling...	416-555-0186	acarter@margiestravel.com	M	70...	1	0.552484051807462	
Mr. Adam J Foster	4524 Ferndale Lane Yaki...	269-555-0137	afoster@woodgrovebank...	M	50...	0	0.691282668500688	
Mr. Adam R Yang	9836 Hanson Lane Berkel...	1 (11) 500 555-0...	ayang@adatum.com	M	80...	1	0.552484051807462	
Mr. Adrian Murphy	1192 Parkway Drive Mill ...	270-555-0127	amurphy@fabrikam.com	M	50...	1	0.621985417835109	
Mr. Adrian A Rogers	5543 Royal Arch Court B...	1 (11) 500 555-0...	arogers@margiestravel.com	M	40...	1	0.621985417835109	
Mr. Aidan B Price	8157 W. Buchanan Both...	269-555-0127	aprice@margiestravel.com	M	40...	1	0.552484051807462	
Mr. Alan Wu	8168 Eaker Way La Jolla ...	913-555-0176	awu@cohovineyard.com	M	30...	0	0.691282668500688	
Mr. Alan Zhao	8981 Carmel Drive W. Lin...	1 (11) 500 555-0...	azhao@fabrikam.com	M	30...	1	0.552484051807462	
Mr. Albert Gutierrez	7656 Ramsey Circle Cha...	984-555-0146	agutierrez@adatum.com	M	70...	1	0.621985417835109	
Mr. Albert Munoz	9371 Corte Del Sol Down...	1 (11) 500 555-0...	amunoz@cpandl.com	M	30...	1	0.621985417835109	
Mr. Albert Ruiz	3397 C Olivera Rd Unit A ...	749-555-0170	aruiz@humongousinsuran...	M	60...	0	0.532371256301275	
Mr. Albert Torres	6513 Beauer Lane Portla...	370-555-0132	atorres@litwareinc.com	M	50...	0	0.691282668500688	
Mr. Albert A Gil	9094 William Reed Dr Lyn...	1 (11) 500 555-0...	agil@fineartschool.net	M	90...	0	0.532371256301275	

- Conditional splits can be used to split the data having different probabilities and only show the customers that are most likely to buy a bike
- The results can be sorted to target those customers with ads etc.

DATA MINING QUERY IN SSIS



FullName	Probability
Ms. Kathryn Tang	0.803423992229237
Mr. Sebastian W Kelly	0.803423992229237
Mr. Shawn A Chande	0.803423992229237
Mr. Levi B Lopez	0.803423992229237
Ms. Melinda Blanco	0.803423992229237
Mr. Jared E Cooper	0.803423992229237
Ms. Margaret K Cai	0.803423992229237
Ms. Mayra Rodriguez	0.803423992229237
Ms. Melody Serrano	0.803423992229237
Mr. Nathaniel C Stewart	0.803423992229237
Mr. Phillip S Sanchez	0.803423992229237
Ms. Mayra C Kapoor	0.803423992229237
Mr. Randy J Cai	0.803423992229237
Mr. Jerome M Alvarez	0.803423992229237
Ms. Gina G Gomez	0.803423992229237
Ms. Cara H Chen	0.803423992229237
Mr. Craig Dominguez	0.803423992229237

APPLICATIONS OF DATA MINING IN BUSINESS

- Establish customer shopping behavior
- Customer retention
- Customer segmentation
- Analyze sales campaign
- Predict future trends
- Increase company revenue
- Fraud detection



REFERENCES

- <https://insightsoftware.com/encyclopedia/data-mining/>
- <https://www.investopedia.com/terms/d/datamining.asp>
- [6 essential steps to the data mining process - BarnRaisers, LLC \(barnraisersllc.com\)](#)
- [Business-Intelligence-using-Data-Mining-techniques-and-Business-Analytics.pdf \(researchgate.net\)](#)
- [What Is Data Mining? | IBM](#)
- [Data Mining in the Business Intelligence \(markovate.com\)](#)
- [What is a Data Warehouse? | IBM](#)
- [Data Mining Query in SSIS \(sqlshack.com\)](#)

