



certinia


How Data Modeling and Machine Learning Optimize Sales Activities for Business Growth.

Jose Santos Salvador & Jose Antonio Ruiz Millan

05/2024

DataOps Team Overview 2024

José Fernández
Data Engineer II




Certinia tenure: 7 years

Background: Software engineer with 16 years of experience dedicated to problem solving with a special focus on backend processes.

Role in Dataops: Technical and architecture, expert on ARR processes and Customer Experience and Operations.

Carmen Biedma
Data Engineer II




Certinia tenure: 3 years

Background: Software engineer with 4 years of experience, Data Science PhD student, experience in customer support and big data.

Role in Dataops: Software engineer with specialization in telemetry, Snowball

Jose Ruiz
Data Engineer I




Certinia tenure: 2 years

Background: Software engineer with 3 years of experience, machine learning PhD student, experience in customer support and data science.

Role in Dataops: Software engineer with knowledge on telemetry, AWS, customer experience and attribution model projects.

José Santos
Data Engineer I




Certinia tenure: 1 year

Background: Software engineer with 2 years of experience in education software development.

Role in Dataops: Software engineer with knowledge of NPS, Productivity metrics, Code Insight and Snowball.

Hicham
Associate Data Engineer

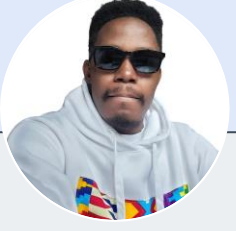


Certinia tenure: Just joined

Background:

Role in Dataops:

Neri Koutchala
Associate DevOps




Certinia tenure: Just joined

Background: associate DevOps engineer with a year of experience, Master's degree in Data science and computer engineering.

Role in Dataops: Cloud platform scalability, efficiency, automations, costs control, logging and monitoring


Vaibhav Goyal
Associate DevOps



Certinia tenure: Just joined

Background:

Role in Dataops: Cloud platform scalability, efficiency, automations, costs control, logging and monitoring




Miguel Muñoz ETL II

Certinia tenure: 2 years

Background: Systems engineer with 14 years of experience and background in system engineering, devops and product management.

Role in Dataops: Project, product and people management. Leads conversations with stakeholders and prioritises the backlog.



Paul Ripley Principal Solution Architect

Certinia tenure: 13 years

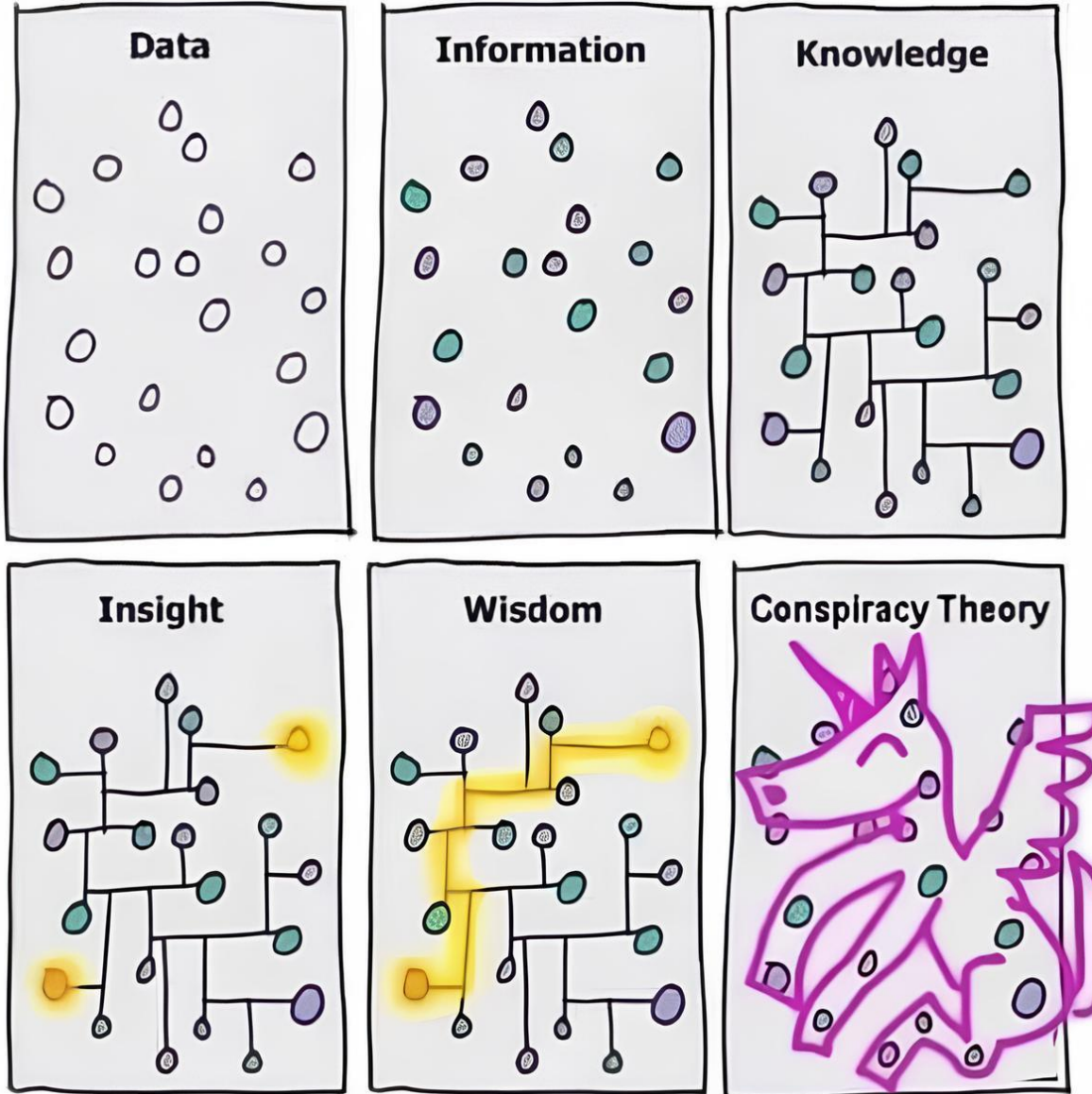
Background: Finance, Sales, Product Management.

Role in Dataops: Technical and architecture, expert on ARR processes, Finance, Sales, Customer Experience and Operations.

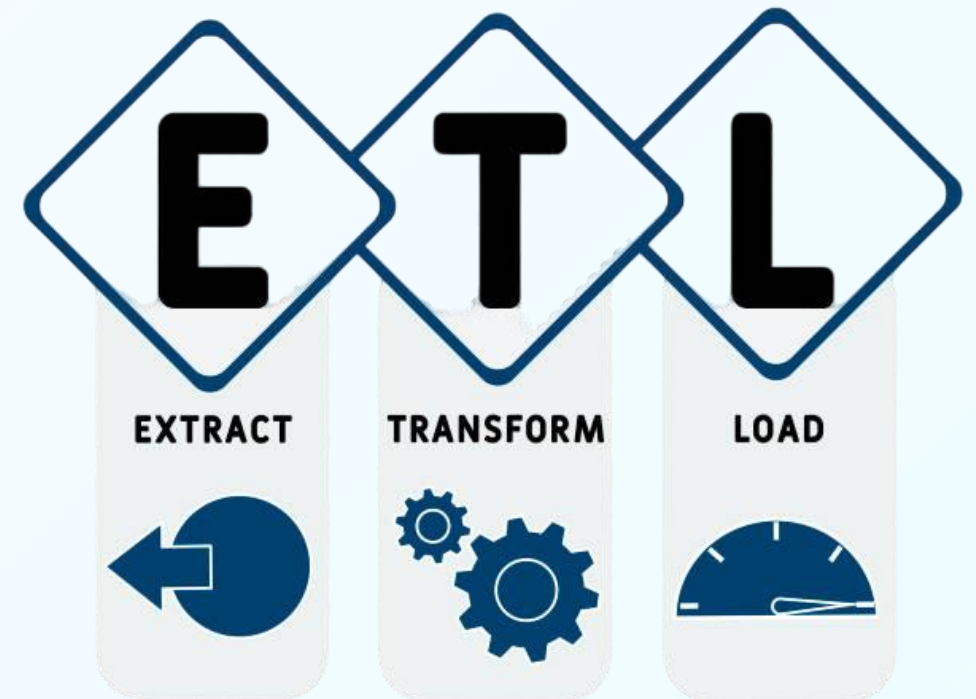
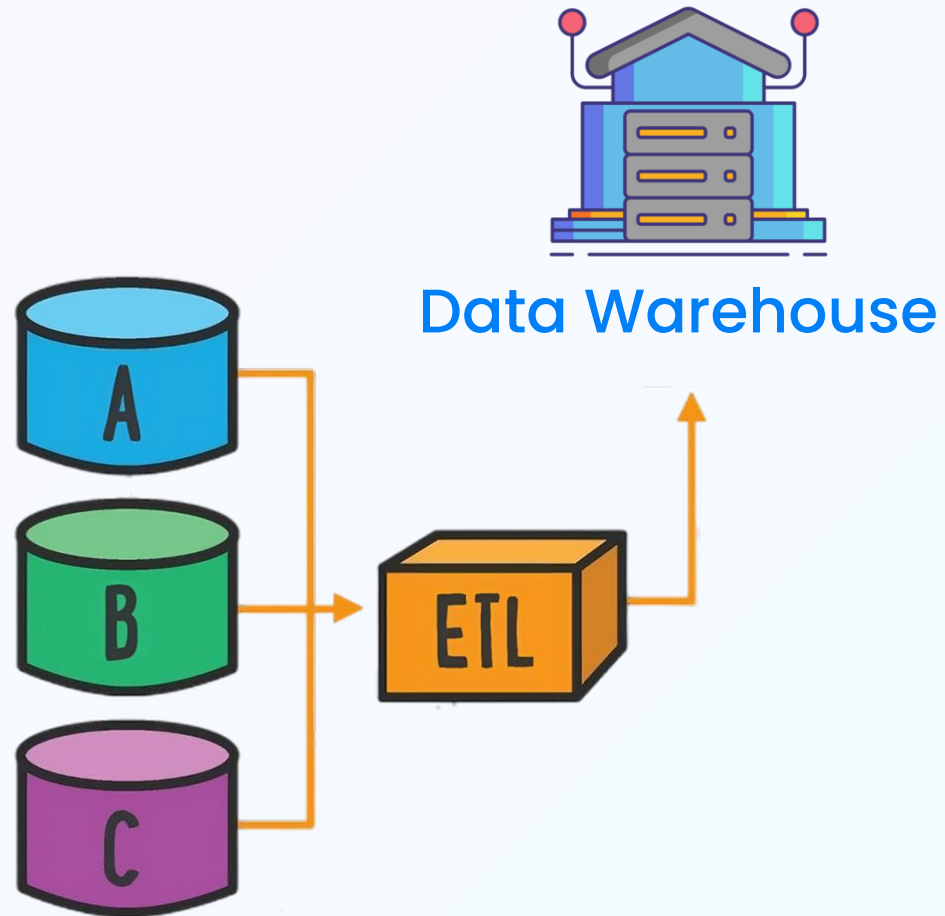
Data modeling



The importance of data modeling



Data modeling: Data Warehouse and ETL



ETL basic example



"José Santos, I'm in"

"José Antonio Ruiz, I'm out"

E

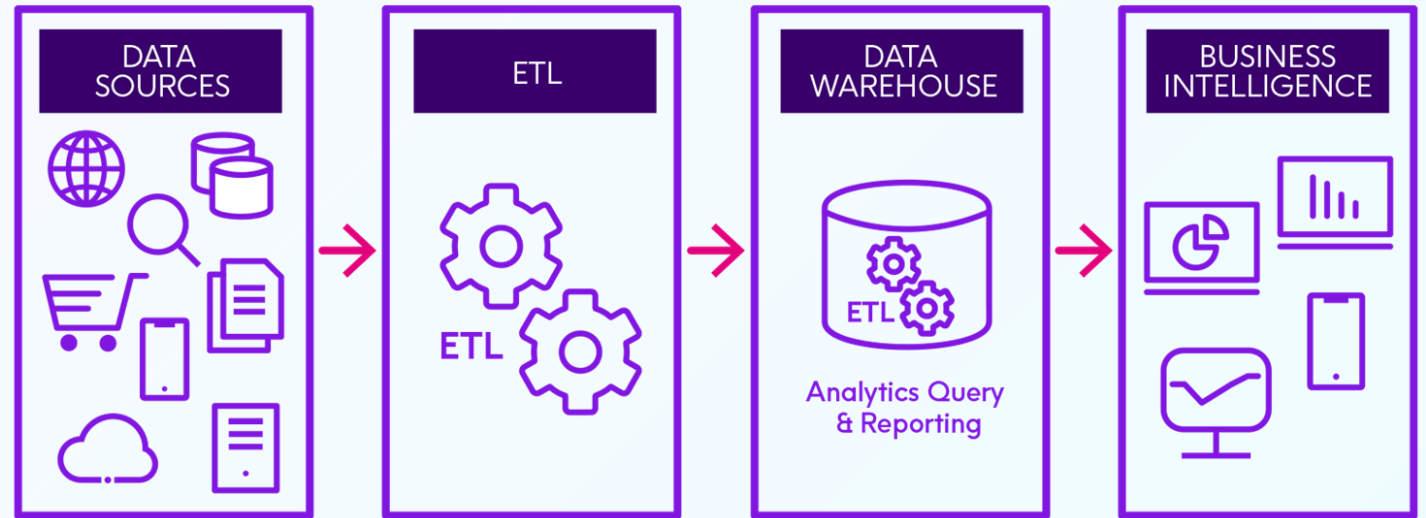
Attendee	Assistance
José Santos	True
José Antonio Ruiz	False

T



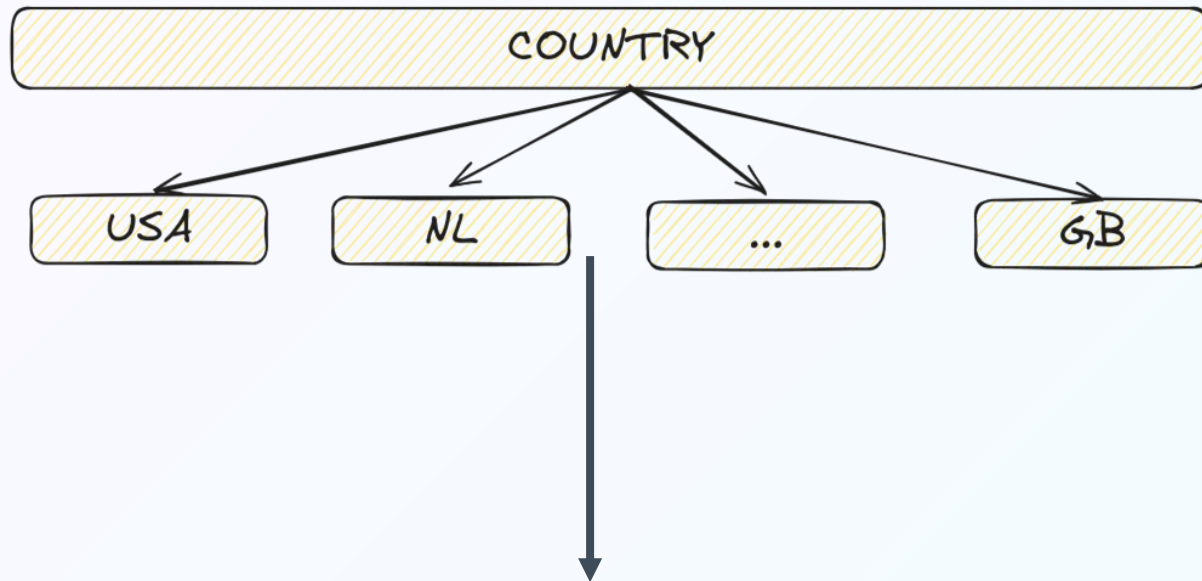
L

Improve ETL process using Polars



Polars lazy API

1



2



```
# Analyzing 27GB data file on 8GB RAM
lf_big = pl.scan_parquet(source='big_file.parquet')

(
    lf_big
    .filter(pl.col("country") == "GB")
    .collect()
)
```


Python and Machine Learning



Python



libraries



Pandas



Numpy



Polars



Tensorflow



Pytorch



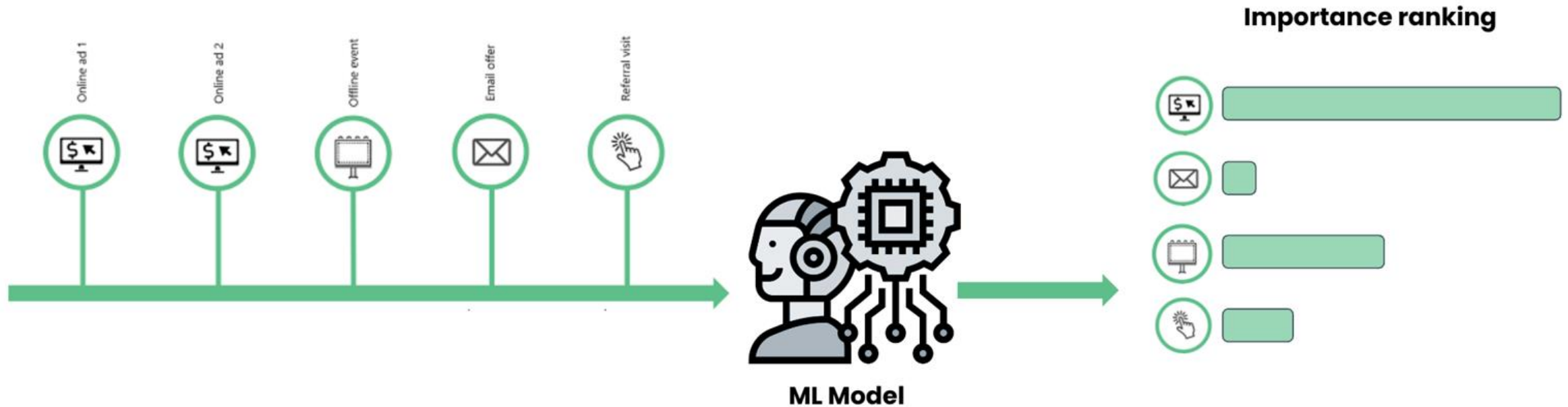
NLTK

spaCy

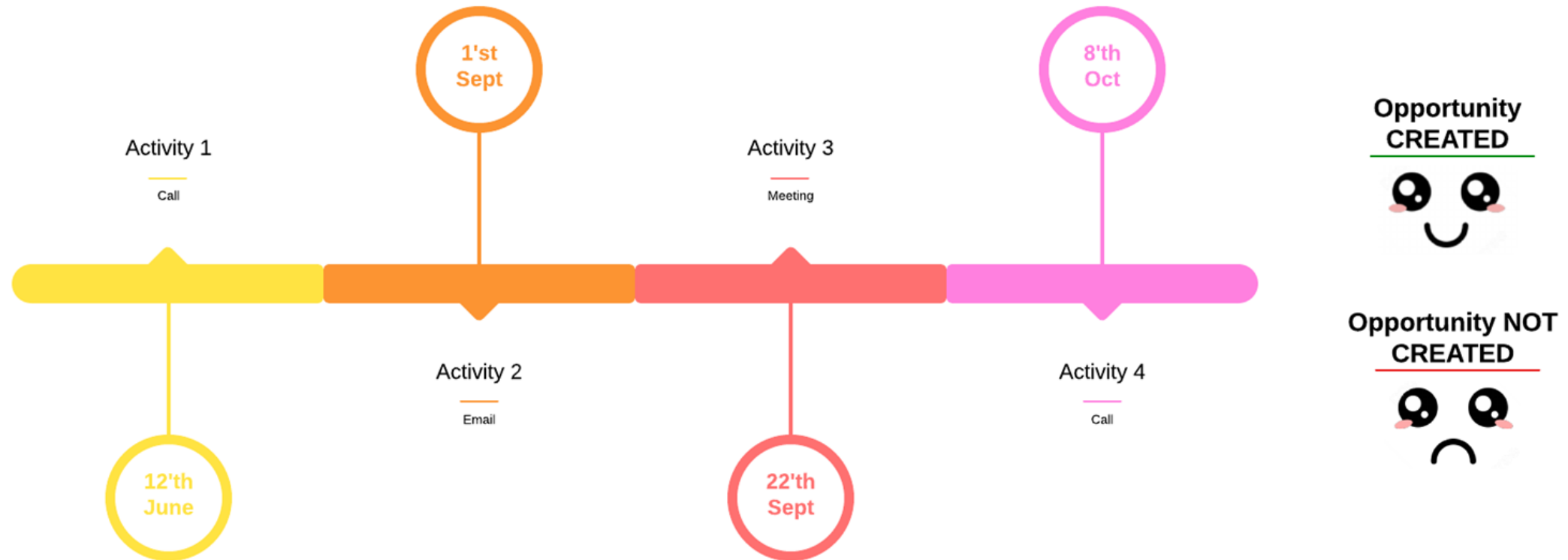


OpenCV

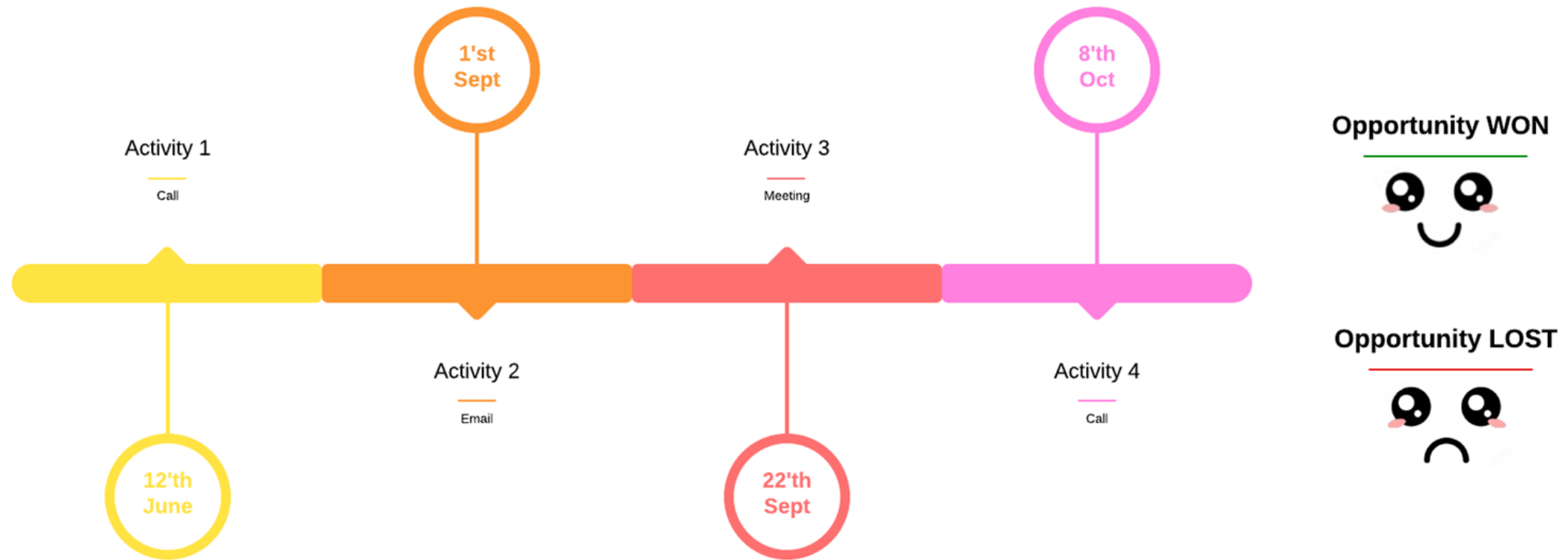
Sales and Marketing challenge



Sales and Marketing first approach



Sales and Marketing final approach



Sales and Marketing final approach

Exploratory data analysis (EDA)

- Statistical analysis
- Correlations
- Feature analysis

Preprocessing data

- Normalization
- Variable creation
- Noise cleaning
- Outliers

Modeling

- [Cross-validation](#)
- [XGBoost](#)
- [Grid-Search](#)

Results presentation

- Metrics (Accuracy and F1 score)
- Matplotlib
- Seaborn

Sales and Marketing results

Metrics

Accuracy

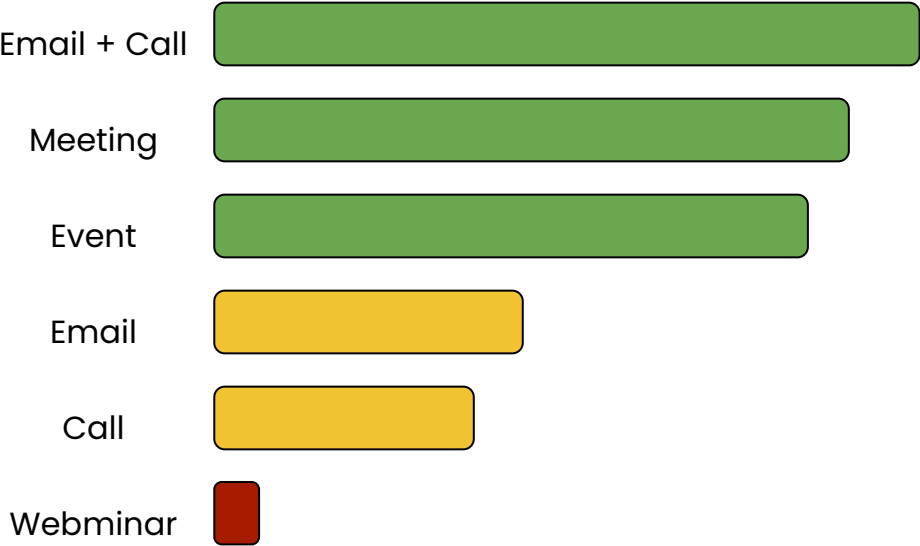
83%

F1 score

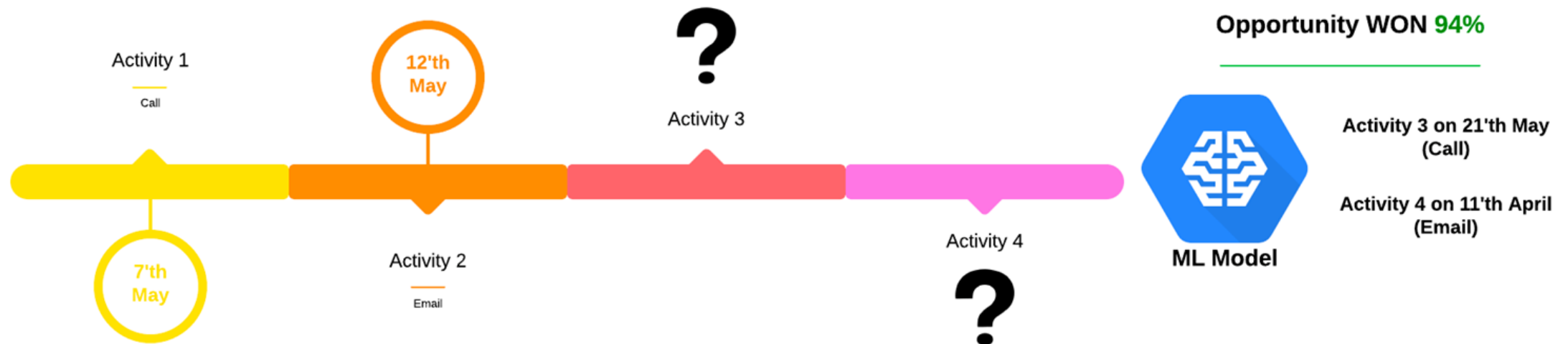
81%

* This is a representation of the actual results, not the actual results..

Ranking



Sales and Marketing future projects



certinia

Services mean business and that means Certinia

