

TUTORING PROJECT MASTER DATA SCIENCE KSCHOOL

BERRIES PRICE DECISION MAKING TOOL

26/06/2020

SUMMARY



Idea



Data & Goals



Techniques



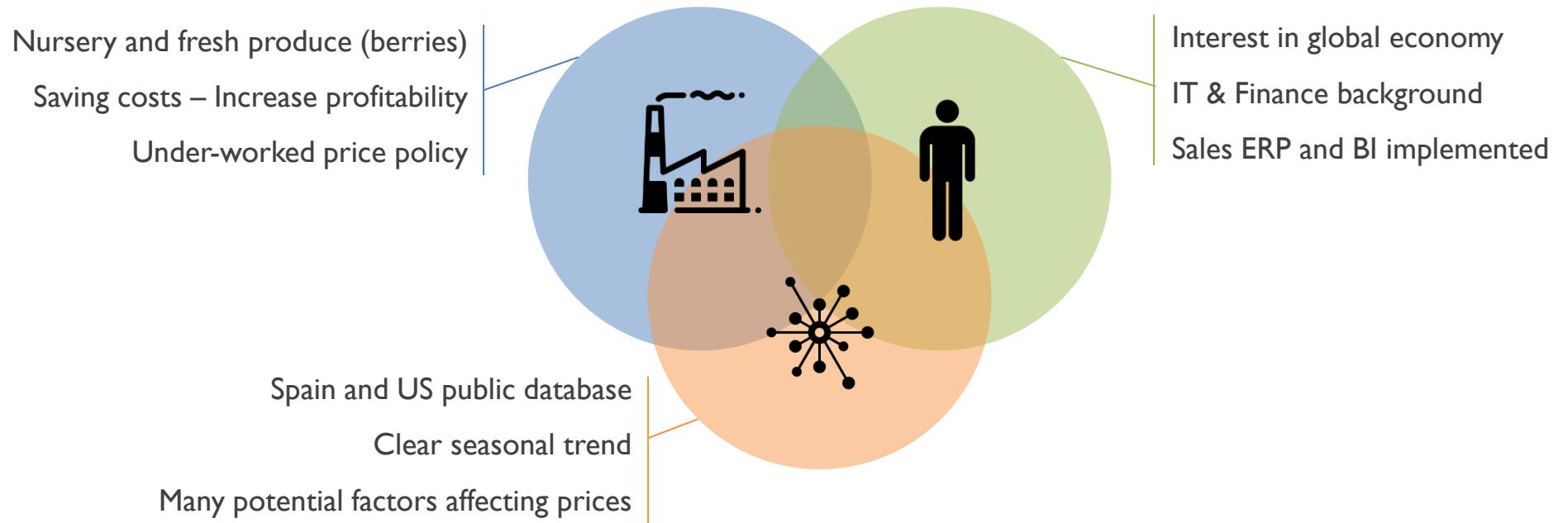
Insights



Planning

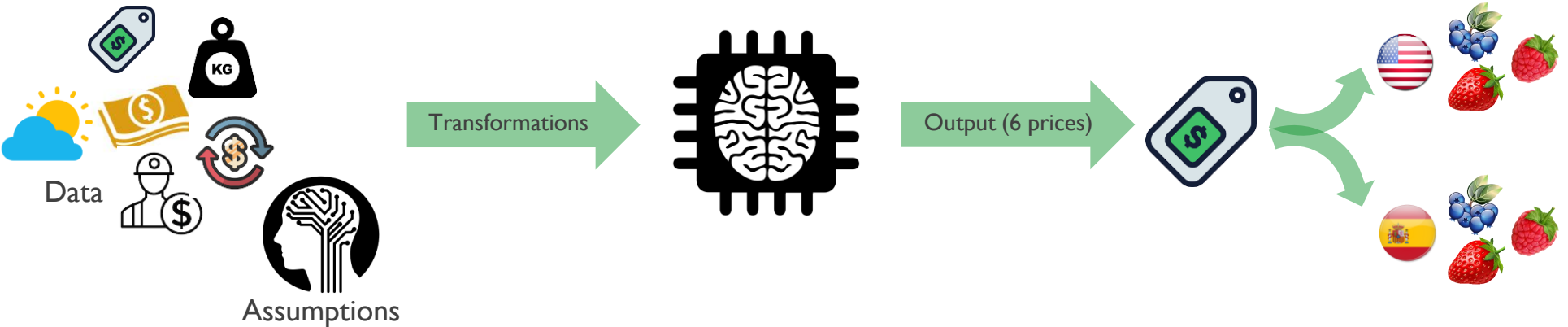
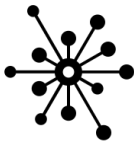


Doubts



Price decision making tool for strawberry, blueberry and raspberry campaigns in Spain and USA

DATA & GOALS



Type	Data Source	Format	Data Kind	Geography	Periodicity	History
External	USDA	URL	Market prices	Mex exports to US	Daily	10 years
External	USDA	URL	Market volumes	Mex exports to US	Daily	10 years
Internal	Sap Business One	Excel	Prices	Mex exports to US	Daily	Pending
Internal	Sap Business One	Excel	Volumes	Mex exports to US	Daily	Pending
External	Junta de Andalucía	Excel	Market prices	Spain	Weekly	4 years
Internal	JD Edwards	SQL ETL	Prices	EMEA	Daily	10 years
Internal	JD Edwards	SQL ETL	Volumes	EMEA	Daily	10 years
External	Min. Agricultura	Excel	Labour Cost	Spain	Yearly	Pending
External	Pending	Pending	Weather conditions	Pending	Pending	Pending
External	BCE	API	Global economy KPI	Europe	Pending	Pending

TECHNIQUES



Ingest

Aritmetic operations
Lagged time-series
SQL Database
Web scrapping
CSV files



Explore

Histogram, boxplot, trend
Correlation matrix
Autocorrelation function



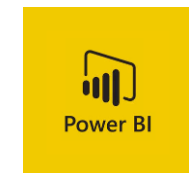
Model

Generalization (supervised)
LM / GLM
ARIMA (SARIMAX)
Decision Trees
Random Forest
Gradient Boosting
XBoost

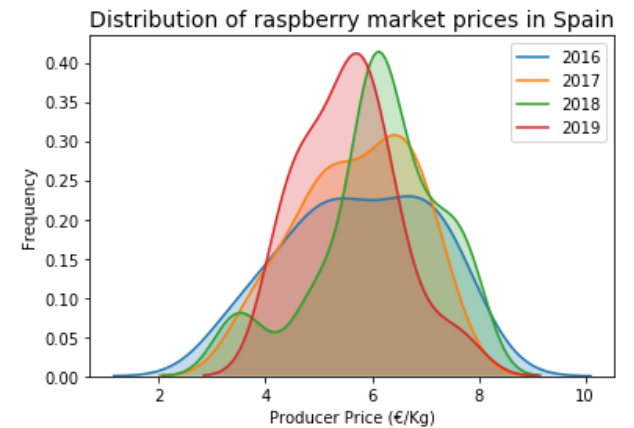
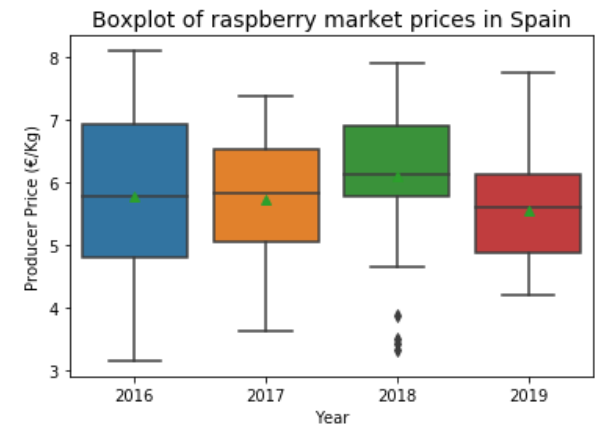
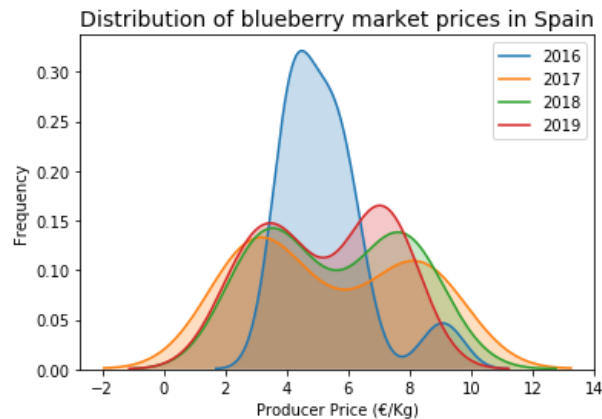
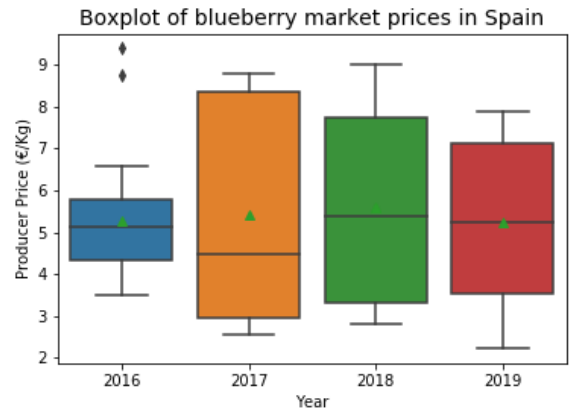
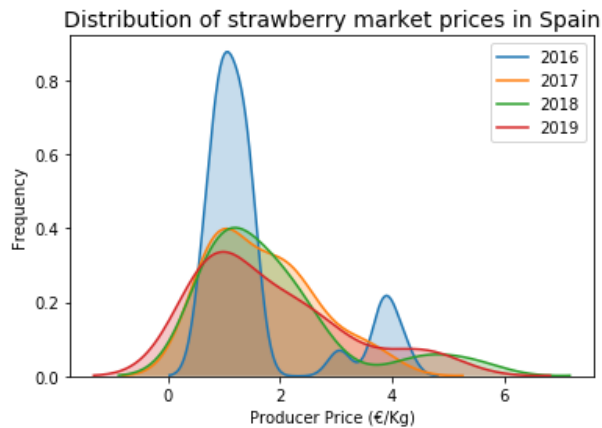
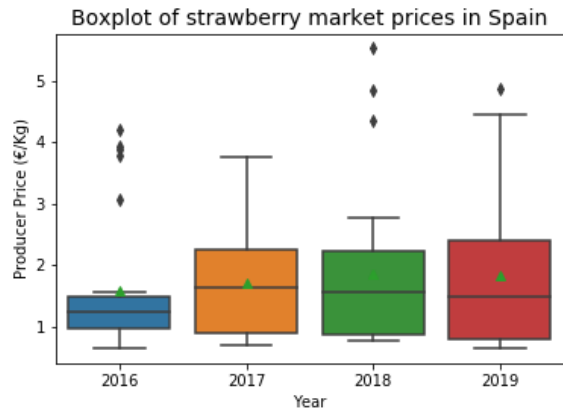


Visualize

Line and multi-line graphs
Histograms
KPI cards



INSIGHTS (SPANISH MARKET)

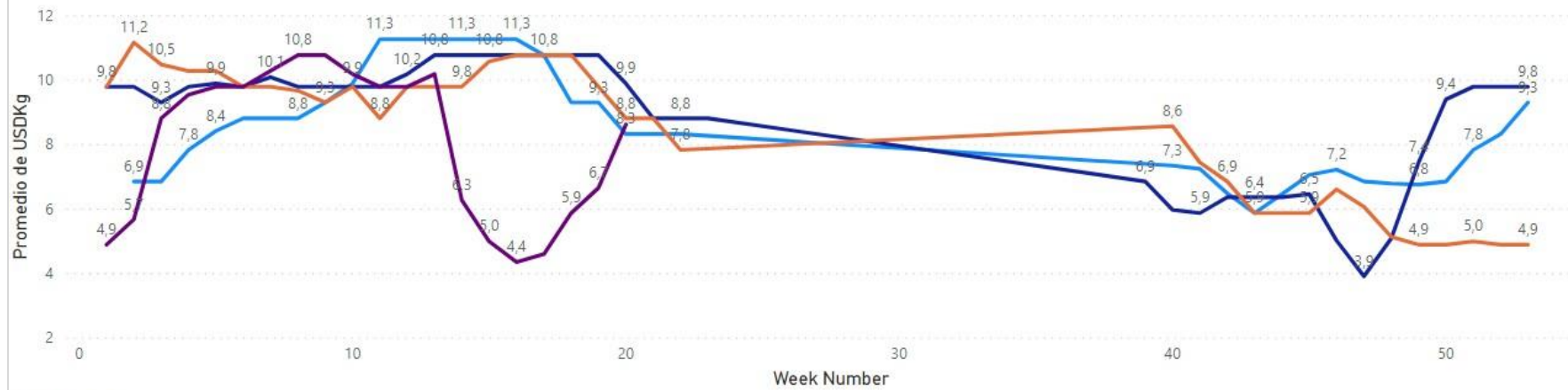


INSIGHTS (USA MARKET – MEX IMPORTS)



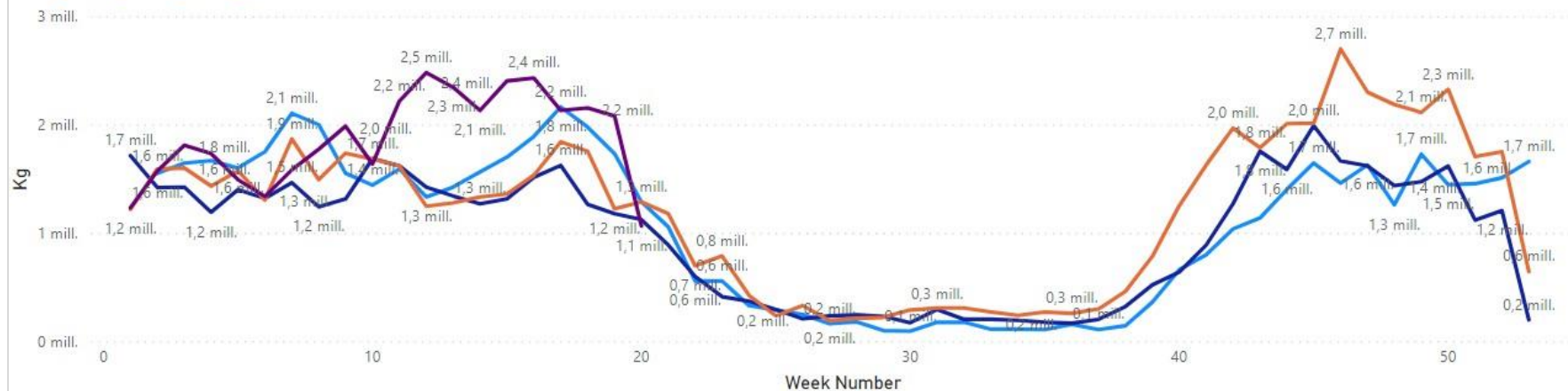
Exports Average Price (USD per Kg)

Year ● 2017 ● 2018 ● 2019 ● 2020



Exports in Kg

Year ● 2017 ● 2018 ● 2019 ● 2020



PLANNING



DUMMY MODEL

ARIMA and simple GLM
with prices and lags

July

MODEL 2.0

Combine Model 1.0 with internal prices
and refine visualization part



June

MODEL 1.0

SARIMAX and other Supervised models
with prices and lags plus labour costs,
weather conditions, other market
prices, global KPIs

August



- Tagged data per crop – region:
 - Spain \approx 500 prices (4 campaigns)
 - US \approx 4500 prices (10 campaigns)
- Meteorological data sources (rain volumes, low temperatures, cloudy weather)
- SARIMAX, GLM, decision trees, random forest
- Power BI vs Website

