Sales Forecasting for Retail in Ecuador

Presented by Kishore Kumar D, Data Science Intern, for Wiseanalytics. This presentation outlines the sales forecasting project. It aims to improve demand forecasting and inventory planning. Understanding external factors is also a key goal.

KK **by Kishore Kumar**

Project Overview

Objective

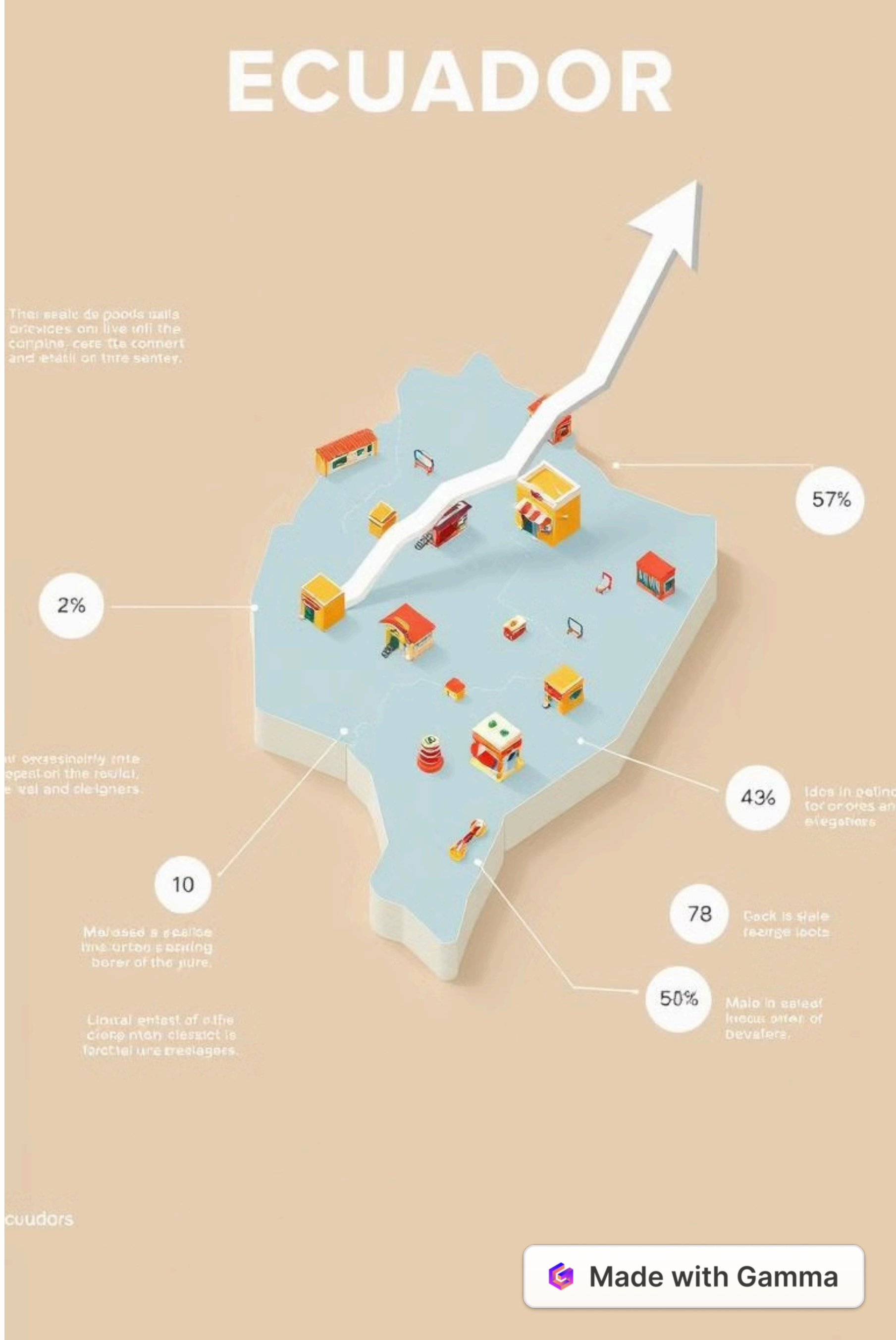
Forecast daily sales for each product family at each store for the next 15 days.

Dataset

Utilize historical sales data, oil prices, store info, and holidays/events.

Business Use Case

Improve demand forecasting, support inventory, promotion planning and understand external factor impacts.



Workflow Breakdown

1

Day 1: Data Processing

Merge datasets and handle missing oil prices using interpolation techniques.

2

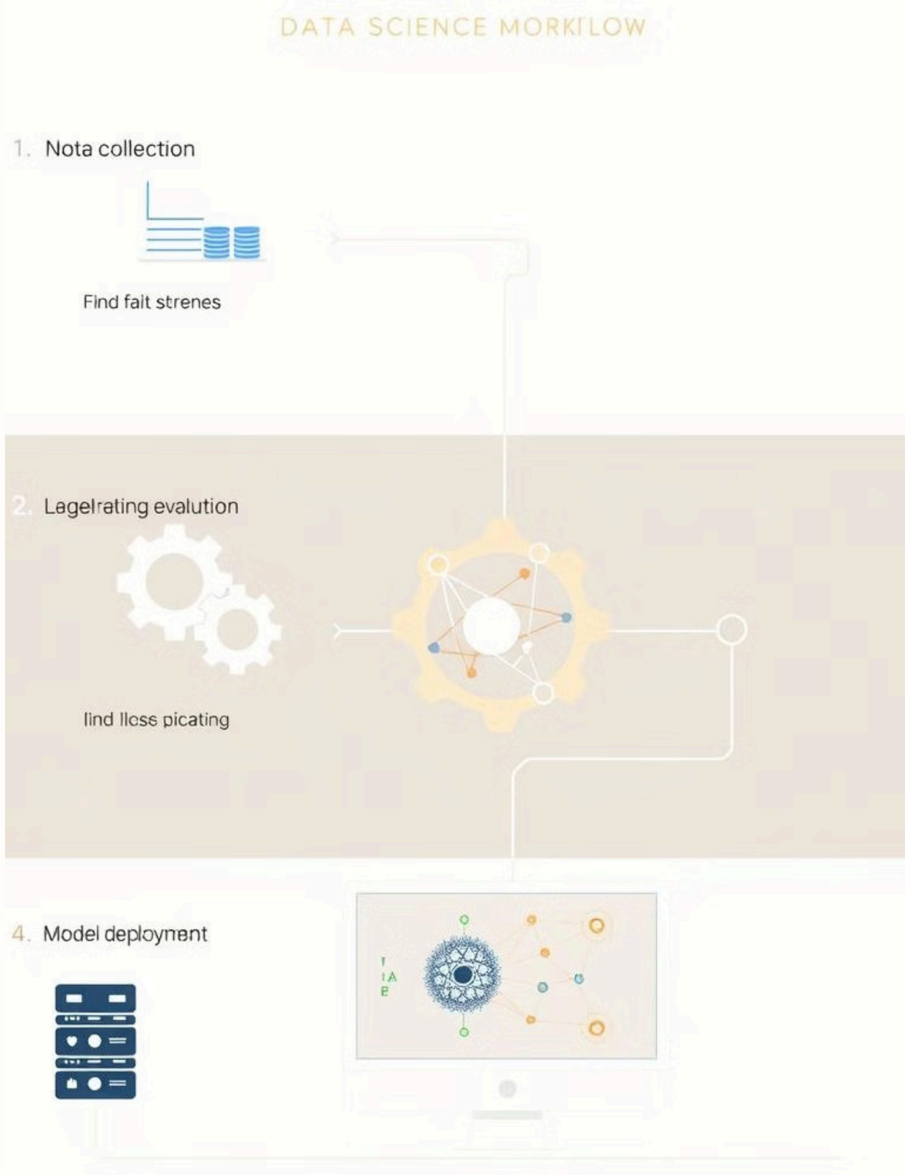
Feature Engineering

Create time and event-based features and build lag and rolling window features.

3

Day 2: Modelling

Train and compare models, visualize results and share insights.



Managie	24500	33.00	34.00	3,360	33.00	3.3.00
Wandian	23500	13.00	28.00	2,000	35.00	3.1.00
Auianzure	49400	13.00	21.00	3.200	23.00	2.0.00
Ancenda	15500	18.00	28.00	2,500	33.00	3,8.00

Data Preprocessing

Missing Values

Imputed missing oil prices using linear interpolation for accurate data.

Merged Datasets

Integrated train, stores, oil, and holidays data for comprehensive analysis.

Added Features

Incorporated payday flags and earthquake impact for enhanced model insights.

Feature Engineering

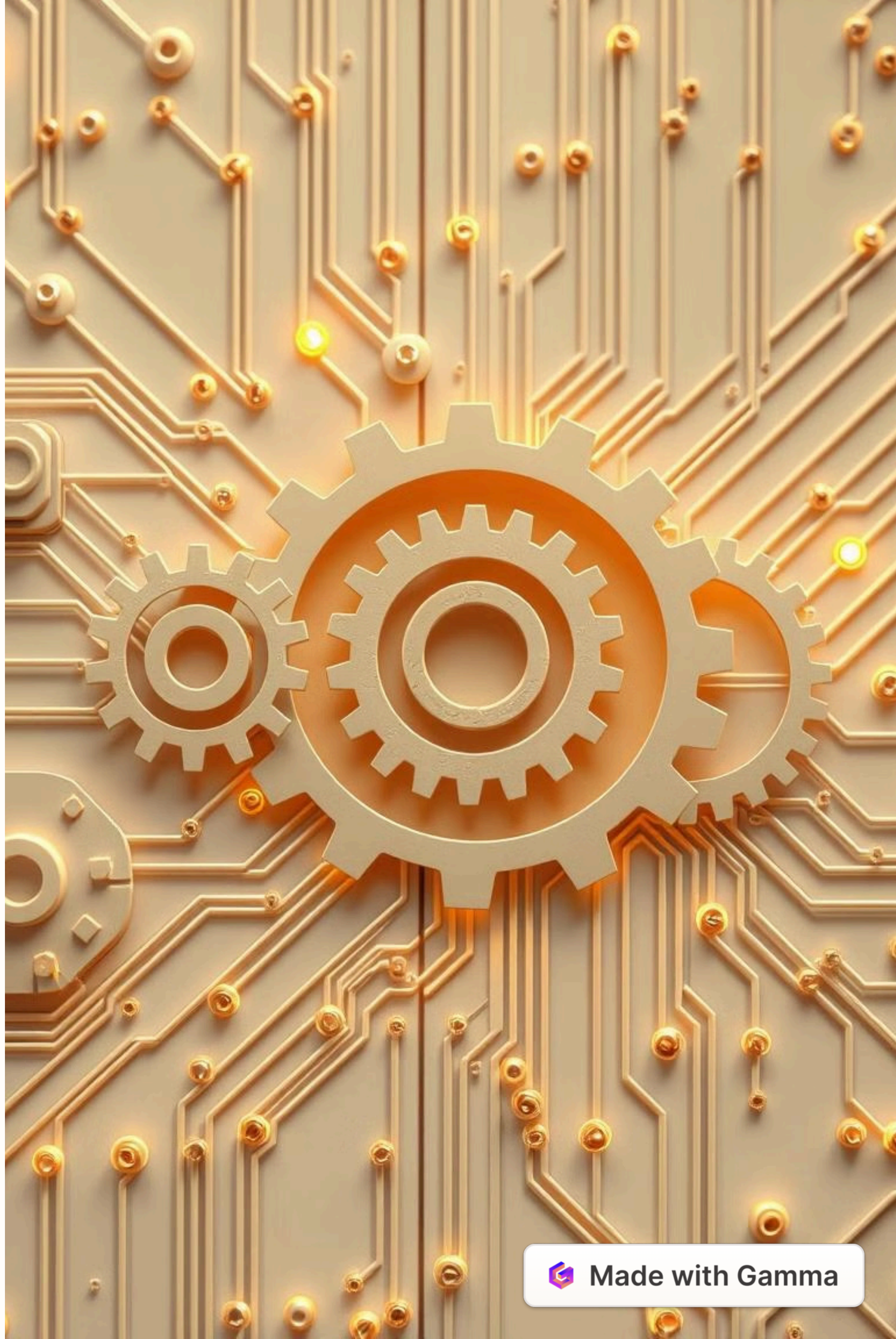
Time Features

Event Features

Lag Features

Rolling Stats

Key features included time-based attributes. Event features are used for holidays. Lag features captured past sales patterns. Rolling statistics smoothed data trends.



Exploratory Data Analysis (EDA)

Sales Trend

Identified seasonal patterns, revealing key sales cycles.

Holiday Impact

Observed higher sales on holidays, impacting inventory.

Oil Correlation

Detected a weak oil price correlation with sales figures.

Model Training

1 Naive Forecast

2 ARIMA

3 Random Forest

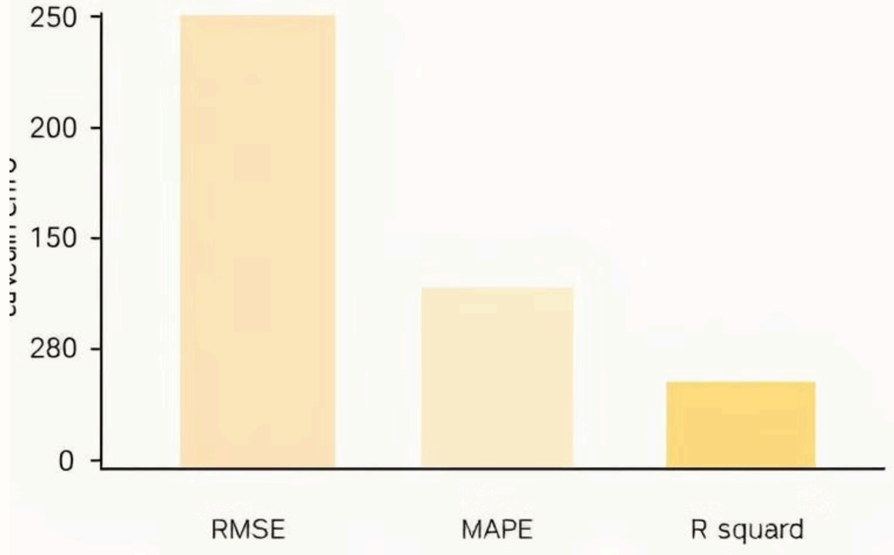
Trained several models on the engineered dataset, including Naive, ARIMA, Random Forest, XGBoost and LSTM. Validation split was used for tuning.

Model Evaluation

LSTM had the lowest RMSE and highest R-squared. LSTM outperformed traditional models in forecasting accuracy.

Model	RMSE	MAPE	R^2
Naive	1740	21.3%	0.62
ARIMA	1587	18.7%	0.69
Random Forest	1395	15.1%	0.76
XGBoost	1260	13.9%	0.82
LSTM	1205	12.8%	0.85

model evaltion metits



Made with Gamma