Siemens AG Lead Time Model

RESEARCH QUESTION:

Can the company use this data to analyse and model lead times in order to:

Optimize supply chain sourcing decisions

Better manage customer delivery expectations

The given data is (Orders, Distance, DeliveryLogs

Research Question

To develop and compare predictive models that estimate the delivery lead time of orders based on various factors (distance, order volume, and destination type),to identify **which model provides the most accurate and reliable predictions.**

Collection & paration

Descriptive Statistics

Modelling (Regression Analysis)

Exploratory D Analysis
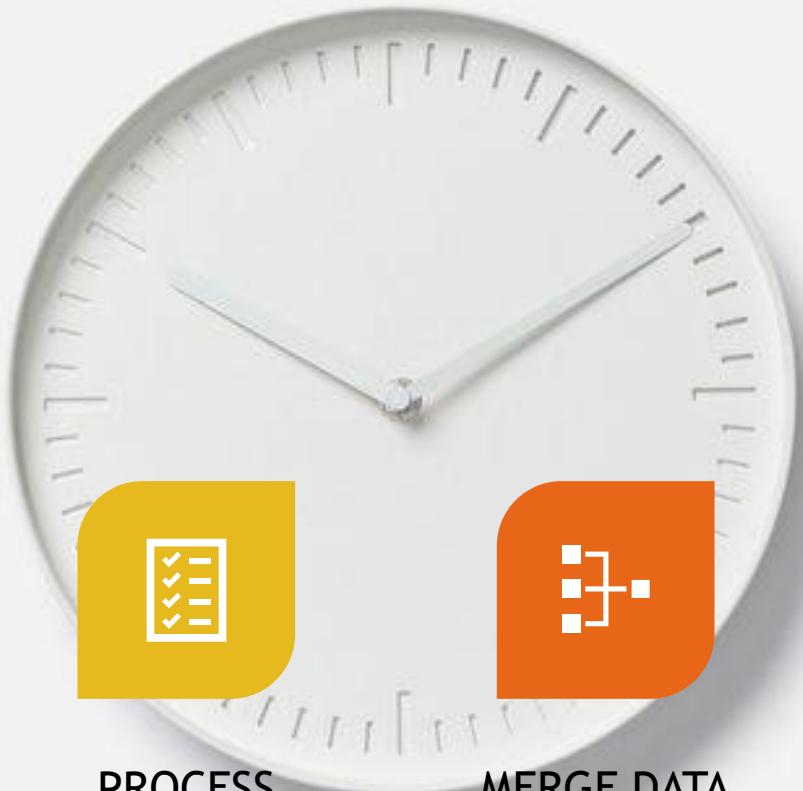
DATA COLLECTION & PREPARATION

PREPARE DISTANCE MATRIX

PROCESS DELIVERY LOGS
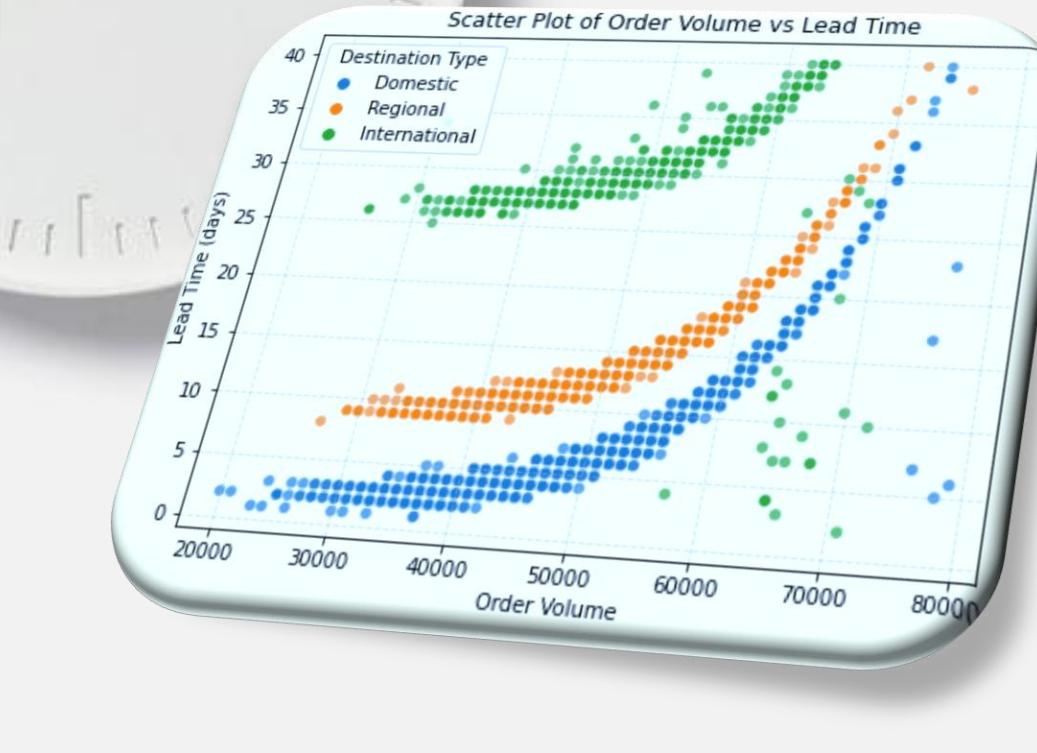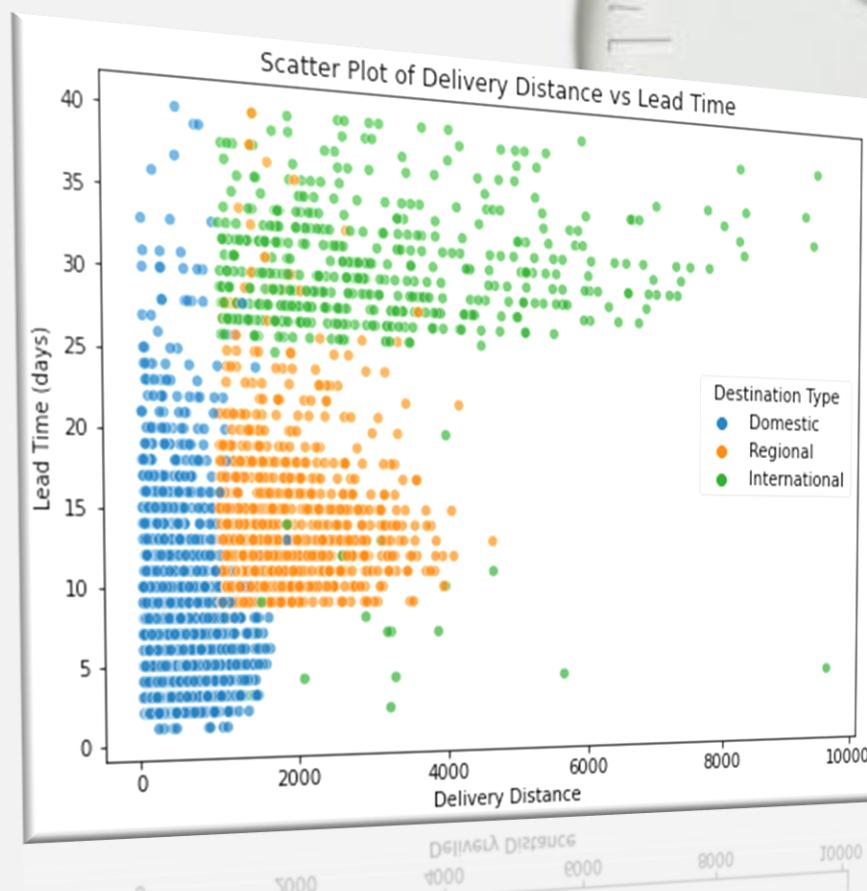
MERGE DATA

SPLIT DATA (70/30)

Exploratory Data Analysis

➤ Which predictor variables provide the most accurate estimation of delivery lead time?

# The Violin Plot for LeadTime and Destination Type

Domestic orders
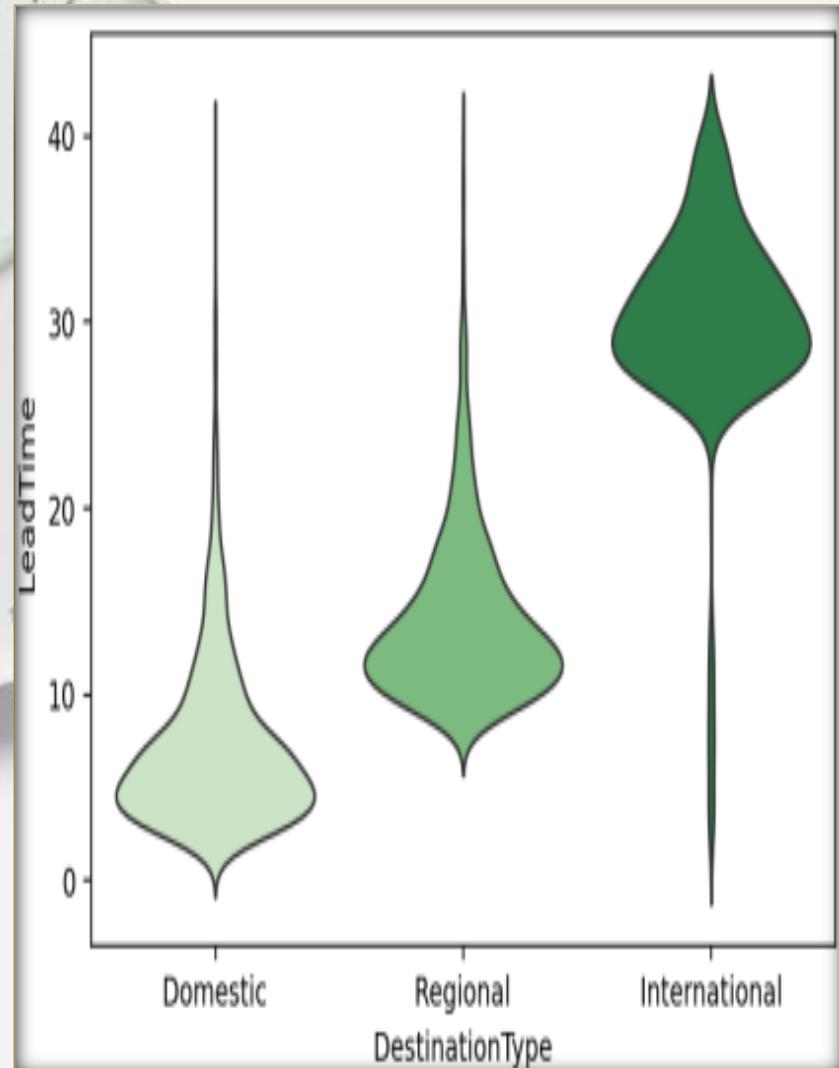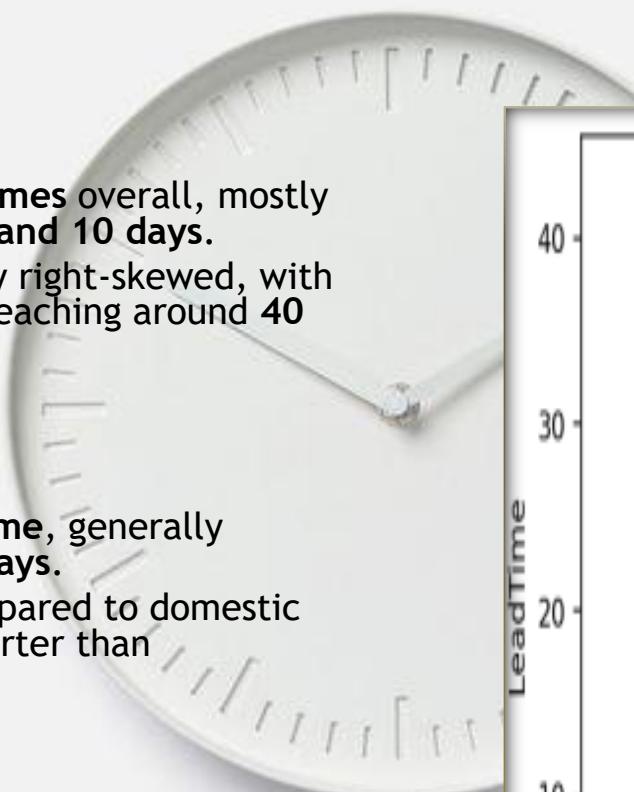
- Have the **shortest lead times** overall, mostly concentrated between **0 and 10 days**.
- The distribution is slightly right-skewed, with a few longer lead times reaching around **40 days**, but these are rare.

Regional orders
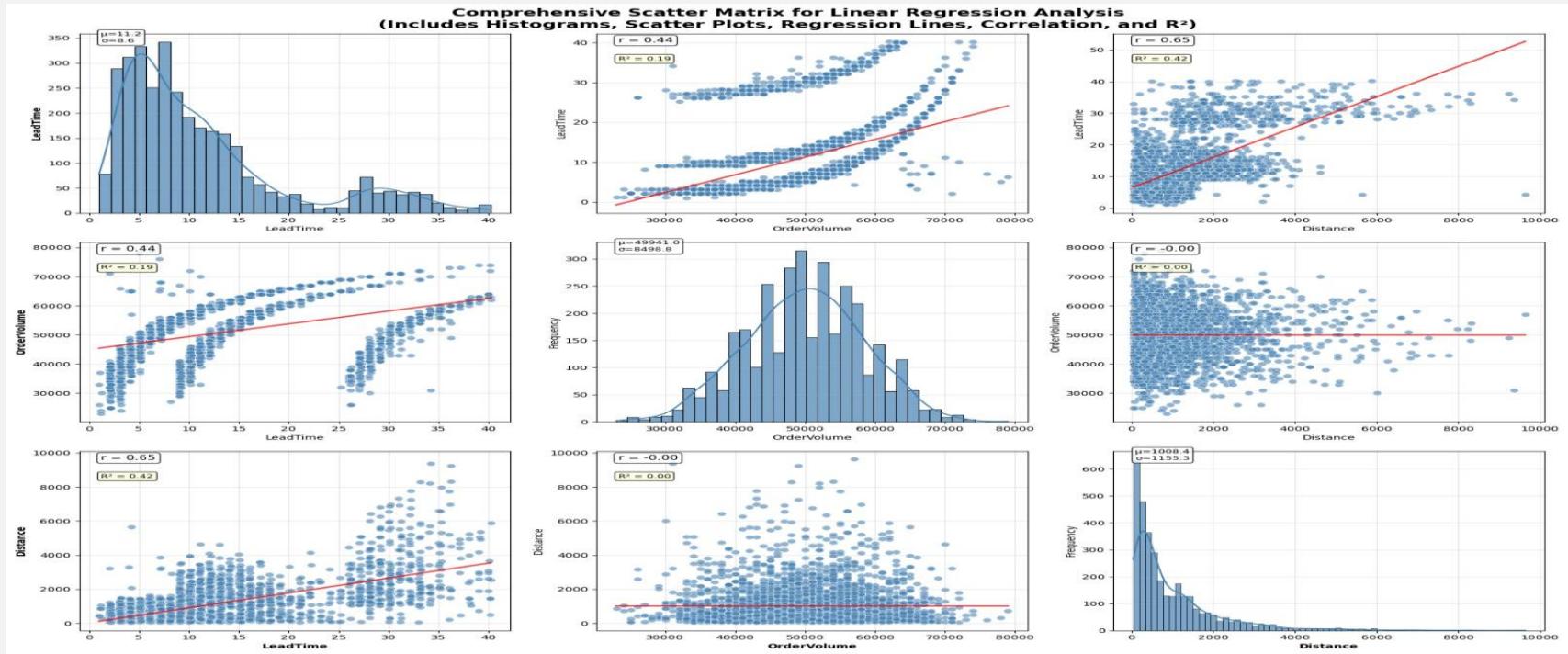
- Show a **moderate lead time**, generally centered around **10-15 days**.
- There's more spread compared to domestic orders, but still much shorter than international.

International orders

- Have the **longest and most variable lead times**.
- The distribution is wider and peaks around **25-30 days**, indicating most international deliveries take significantly longer.
- There are also a few very short and very long lead times, showing greater variability.

Correlation Analysis

Comprehensive Scatter Matrix for Linear Regression Analysis
(Includes Histograms, Scatter Plots, Regression Lines, Correlation, and R^2)

Relationships Between Key Predictors and the LeadTime
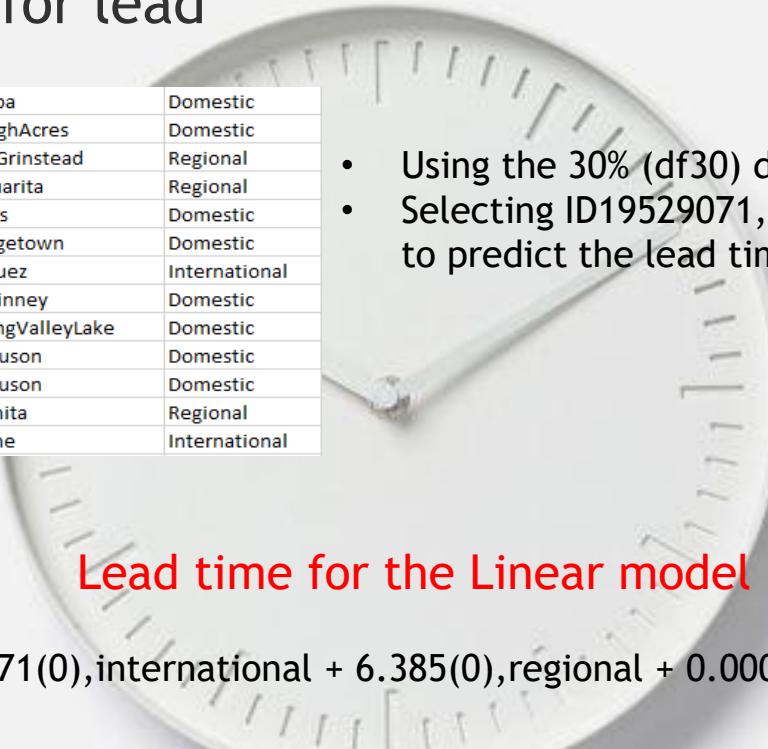
- LeadTime is positively influenced by both OrderVolume and Distance, but Distance has the strongest effect.
- OrderVolume and Distance are independent variables (They are not linearly related)
- In logistics , longer distances and larger orders generally require more lead time, but distance plays a more dominant role.
- other factors not shown here (like the **DestinationType**- Domestic, Regional, International), as these are likely to be important, especially for explaining the variation in lead times for similar distances.

Model Development

➤ Predictive models for lead time.

4422	ID54661765	51000	Miramar	Kimba	Domestic
4423	ID35991875	55000	WestDesMoines	LehighAcres	Domestic
4424	ID57589180	39000	Broomfield	EastGrinstead	Regional
4425	ID70513181	57000	WestPark	Sahuarita	Regional
4426	ID50040803	49000	SouthLakeTahoe	Jenks	Domestic
4427	ID88151736	52000	Allendale	Bridgetown	Domestic
4428	ID33987245	37000	CarneysPoint	Allouez	International
4429	ID19529071	42000	FiveCorners	McKinney	Domestic
4430	ID25378801	53000	Nolensville	SpringValleyLake	Domestic
4431	ID11701119	42000	Shiremoor	Ferguson	Domestic
4432	ID35046287	41000	Hockley	Ferguson	Domestic
4433	ID80715382	45000	Miramar	Wichita	Regional
4434	ID53129530	47000	Weymouth	Blaine	International

- Using the 30% (df30) dataset to test the model:
- Selecting ID19529071, ID53129530 and ID70513181 to predict the lead time
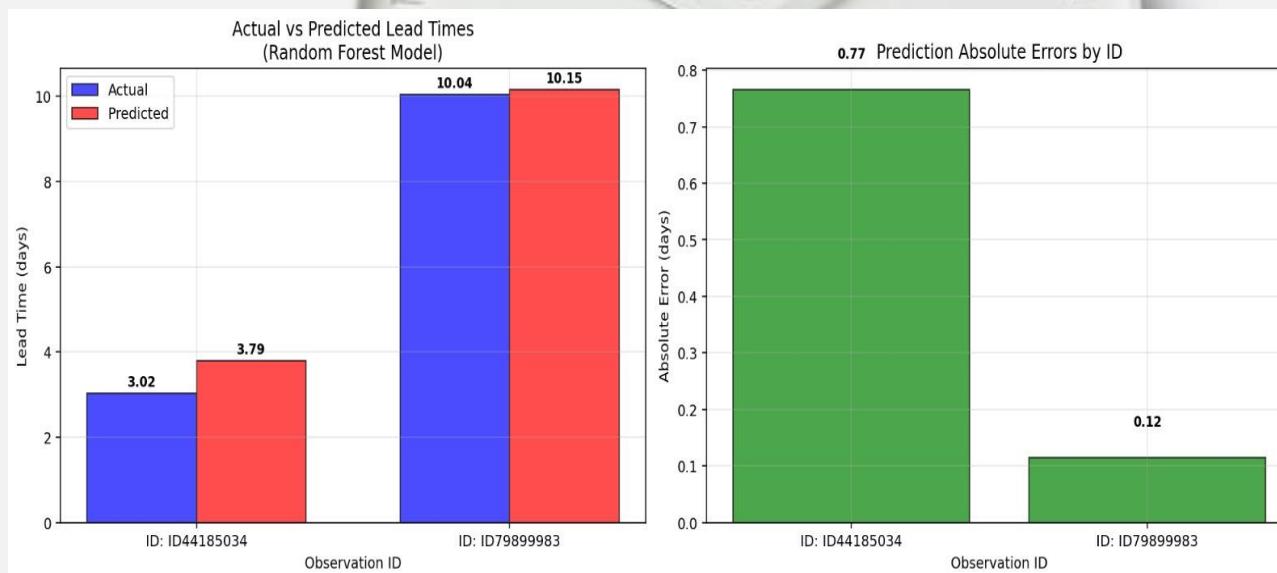
Lead time for the Linear model

- Y,fitted = -14.7845 + 22.2971(0),international + 6.385(0),regional + 0.0004(42000) + 0.0002(80.1) = 2.03 days
- Y = 4.12 days
- Y,fitted = -14.7845 + 22.2971(1),international + 6.385(0),regional + 0.0004(42000) + 0.0002(80.1) = 27.20 days
- Y = 29.17 days
- Y,fitted = -14.7845 + 22.2971(0),international + 6.385(1),regional + 0.0004(57000) + 0.0002(1520.9) = 14.70 days
- Y = 16.04

- For International Purchases the residuals will be low and for domestic purchases the residuals will be a bit higher than international Purchases while regional Purchases will also have high residuals too

Lead time for the Random Forest
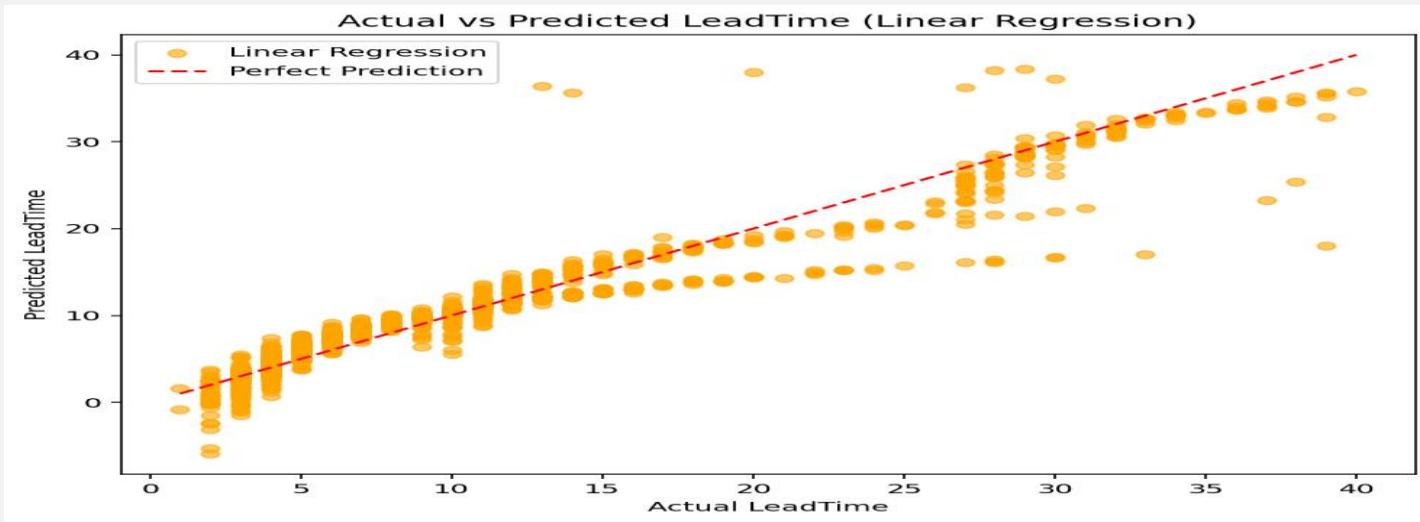
3341	ID43664762	31000	Cirencester	Kimba	International	
3342	ID44185034	41000	Jenks	Shiremoor	Domestic	
3343	ID79899983	41000	EllicottCity	California	Regional	
3344	ID28843816	55000	Sahuarita	McKinney	Regional	
3345	ID43104829	54000	Broomfield	BelAirNorth	Domestic	
3346	ID60945973	56000	Repentigny	WestPark	Domestic	
3347	ID11055049	52000	Simcoe	Bessemer	Domestic	
3348	ID37814319	38000	Upminster	California	Domestic	
3349	ID14988714	47000	Seattle	Bridgetown	Regional	

- Using the 30% (df30) dataset to test the model:
- Selecting ID19529071, ID53129530 and ID70513181 to predict the lead time

Actual vs Predicted Lead Times (Random Forest Model) / 0.77 Prediction Absolute Errors by ID
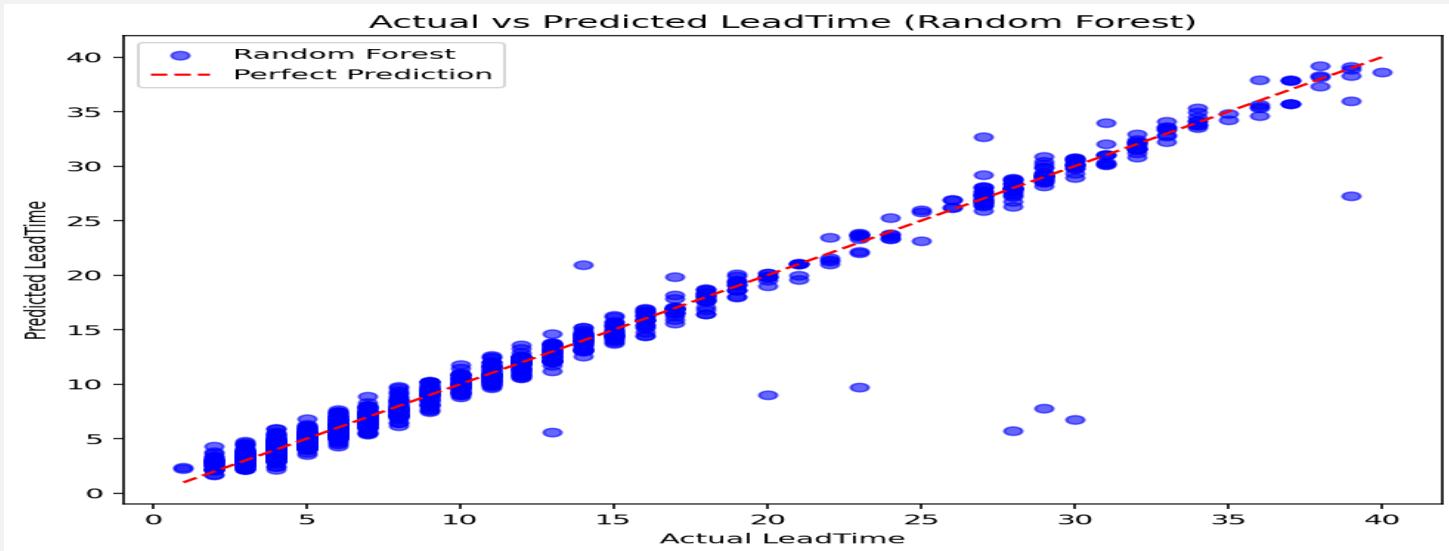
- **Observation ID44185034:** The model's prediction was extremely close to the actual value. It overestimated the lead time by only **0.11 days** (about 2.6 hours). This is a very accurate prediction.
- The model's prediction was less accurate for this observation. It underestimated the lead time by **0.77 days** (about 18.5 hours). This is a more significant error, though not necessarily catastrophic depending on the business context.

Results Visualization


Actual vs Predicted LeadTime (Linear Regression)
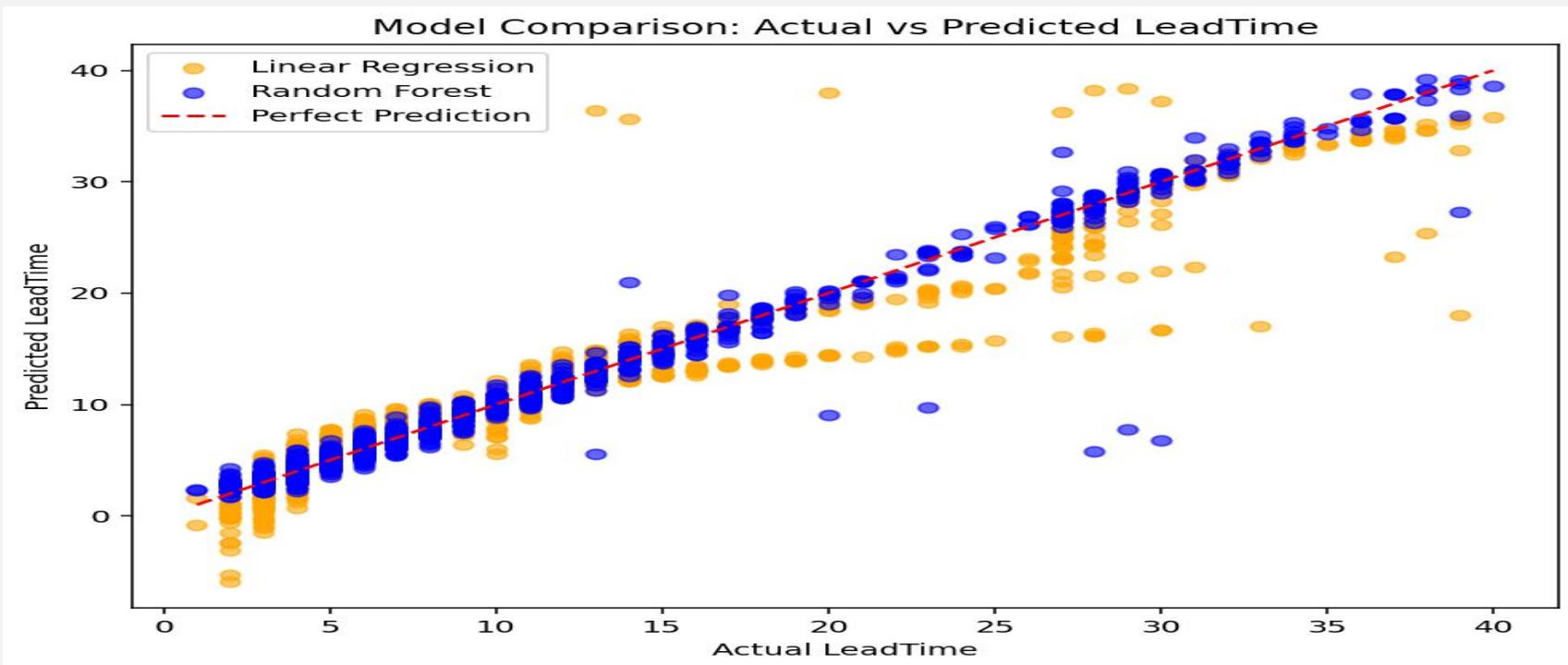
- The Linear Regression model shows **strong but imperfect predictive performance**. Predictions align closely with the perfect prediction line at lower LeadTimes but deviate more for larger values. This suggests that while the model captures the general linear relationship well.
- The linear Regression has a **R² of 0.9214** indicating that **over 92%** of the variation in LeadTime is predicted accurately by the model.
- The model has an MSE of **5.63** and MAPE (Mean Average Percentage Error) of **19.08%** indicate moderate prediction errors, showing that while the model captures the general trend accurately, there is some deviation at higher LeadTimes. Overall, the model performs very well for a linear approach.
- While the model has a MAPE of **19.08%** this shows that on average, the model's predictions are **off by 19.08%** from the true LeadTime this is good but not ideal accuracy
- This is a **Good Model**.

Results Visualization

Actual vs Predicted LeadTime (Random Forest)

- The Random Forest model provides an almost perfect prediction of LeadTime. The predicted and actual values show an exceptionally close match, with most data points lying directly on the perfect prediction line.
- Errors are minimal and randomly distributed, indicating that the model captures both simple and complex relationships without systematic bias.
- Compared to a perfect predictor, the model performs outstandingly well, achieving near-ideal accuracy and reliable predictions.
- The Random Forest Model has a **R² of 0.9736** indicating that **over 97%** of the variation in LeadTime is predicted accurately by the model.
- Predictions are very close to actual values which indicates a low average error and MAPE (Mean Average Percentage Error) of **8.65%,** Since the MAPE is below **10%** the model is considered **very accurate** in most regression tasks.
- The model fits the data extremely well and produces highly accurate predictions and the predictions align closely with the "perfect prediction" line.
- This indicates that the model captures the underlying relationship between Predictor variables and *LeadTime* effectively, with little unexplained variance.

Conclusion

Model Comparison: Actual vs Predicted LeadTime

The Random Forest model should be selected for deployment. It effectively captures the underlying patterns in the data, leading to reliable and accurate lead time forecasts.

By feeding the model new order details (Volume, Distance , Destination-Type).
The Linear Regression model is not suitable for this specific task.

Thus we can conclude that this data-driven approach will directly enable the company to optimize its sourcing strategies based on reliable delivery performance data and expertly manage customer expectations, leading to a more efficient, reliable, and customer-centric supply chain

Thank You

Presented by: [Your Group Name]

Insert Chart or Visualization Here