

Company A Logo

# Improving the Delivery of Machine Learning Insights


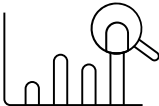
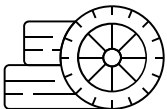
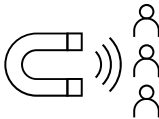


Data Science in the Aftermarket Industry

Intern name  
DSBA-6400  
April 29, 20xx



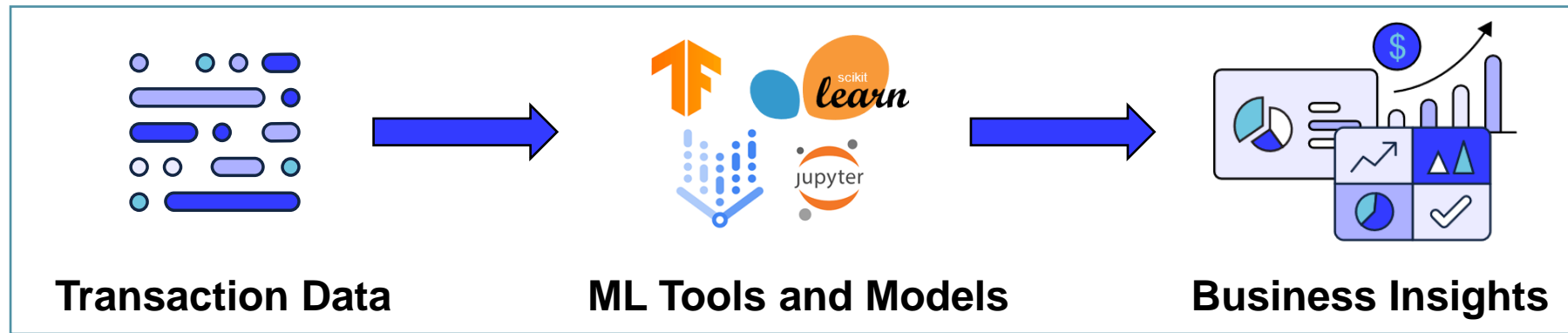
# Company A Drives Better Decisions for the Aftermarket Industry

Transaction Data → Analytics → Business Insights

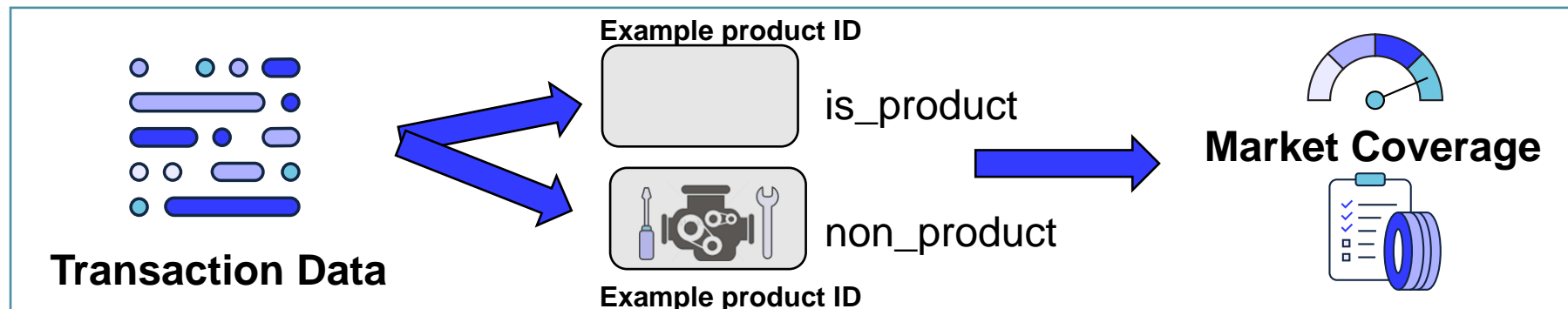
Productivity + Profitability	Marketing + Sales	Vendor Automation
<div></div> <div><b>Profit Optimization</b> Maximize profit by leveraging backend rebates and benefits</div>	<div></div> <div><b>Pricing Insights</b> Gain full access to market data &amp; analysis on local, national, and e-commerce product prices</div>	<div><b>Auto-Replenishment</b> Automatically maintain your SKU-level inventory and take the worry out of balancing supply and demand</div>
<div></div> <div><b>Inventory Optimization</b> Optimize SKU mix for each location based on local demand</div>	<div></div> <div><b>Lead Generation</b> Increase store traffic and sales through a lead generation tool leveraging customer data</div>	<div><b>Inventory Visibility</b> Improve customer satisfaction by locating the right product at the closest location.</div>
<div></div> <div><b>Market Intel Planning</b> Enhance sales strategies based on market signals and actionable insights provided in an easy-to-use dashboard</div>	<div></div> <div><b>Program Optimization</b> Increase performance of rebates, promotions, and campaigns with actionable insights</div>	

# Machine Learning is a Key Driver of Business Insights

- Company A employs multiple, varied ML models to deliver insights on transaction data



- The 'is\_product' classification model separates products from other transactions

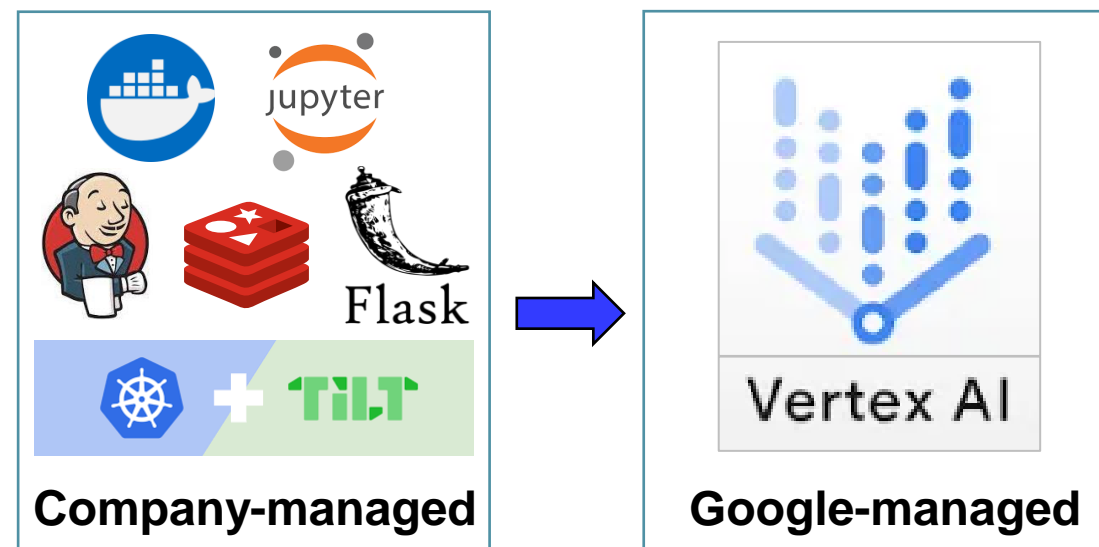


# Growing Demand Requires Scaling of ML Delivery through MLOps

- Company growth & product market changes throughout 20xx were substantial
- **Business Case:**
  - 1) Growth brings more demand for ML, requiring scalable MLOps
  - 2) Big changes in data requires reevaluation of current insights

## Project Objective

1) Build a **machine learning pipeline** for the `is\_product` model within Google Vertex AI, while 2) concurrently **evaluating** and **improving performance** both in terms of processing efficiency and predictive accuracy.

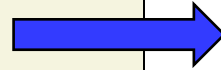




# Business Value Drives Task Development

## Business value propositions

1. **Confirming the viability** of Vertex platform also enables other parallel efforts to begin
2. Prioritizing the ongoing infrastructure changes **reduces total work required**
3. Prioritizing improvement provides the **most value** to the largest audience
4. Vertex pipelines **support Analytics team** to scale MLOps as the company grows into future demand



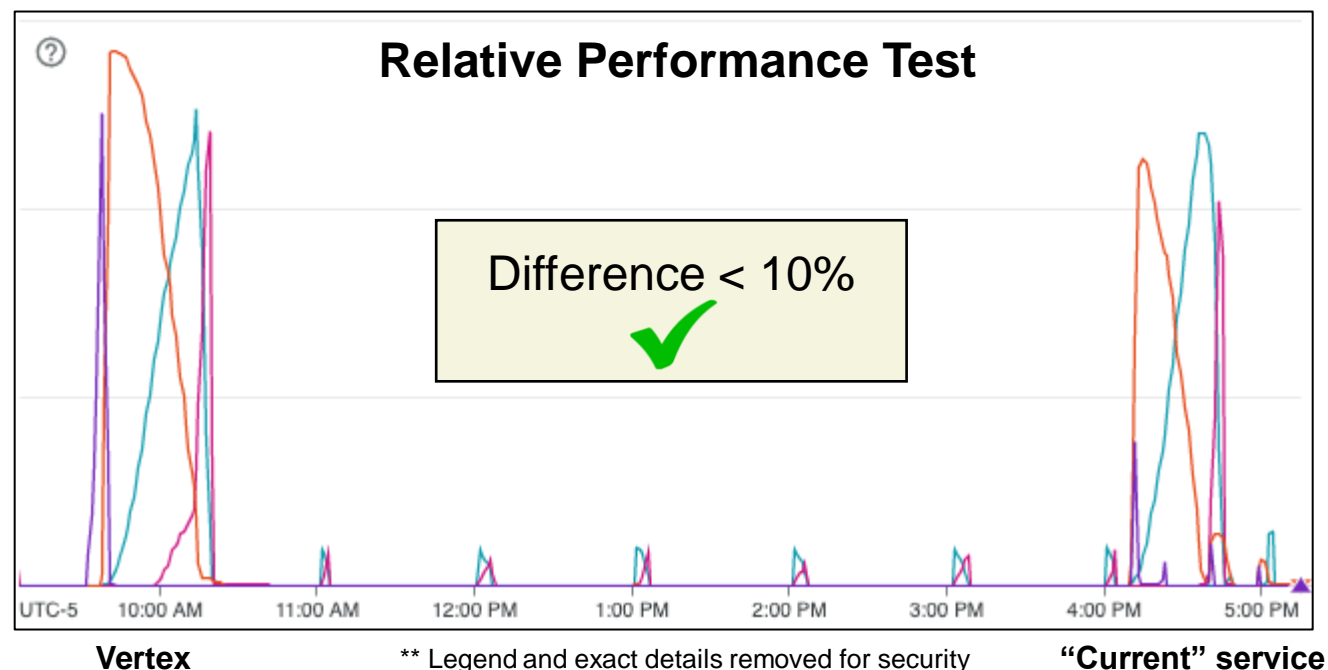
## Key tasks developed

1. Move `is\_product` model from company-managed endpoint to a Vertex endpoint and **validate performance**
2. Replace and **reproduce the company-managed service** that previously hosted the model
3. Evaluate and explore options for **improving `is\_product`** performance
4. Implement the complete **pipeline** in Vertex

# 1. Vertex is a Viable Option for Hosting ML Models

## Key insights to support parallel and future projects

1. Vertex endpoints incur a small standing cost
2. Auto-scaling is effective, but slower than expected
3. Recommend to use HTTP requests—much faster than Google's software development kit (SDK) for python



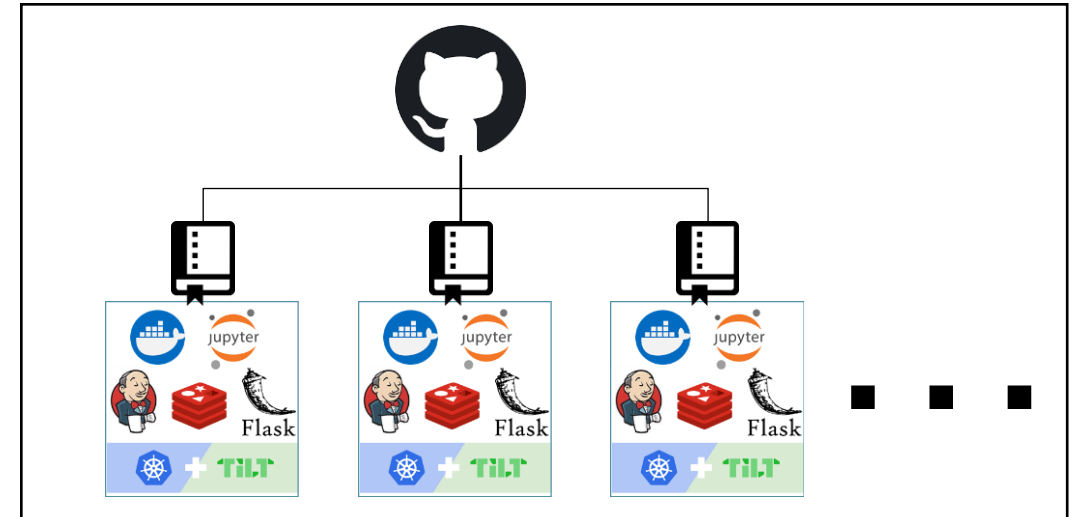
## 2. Replacing the Old Service Reduces Overall Work

### Google Cloud Function (GCF) as a solution

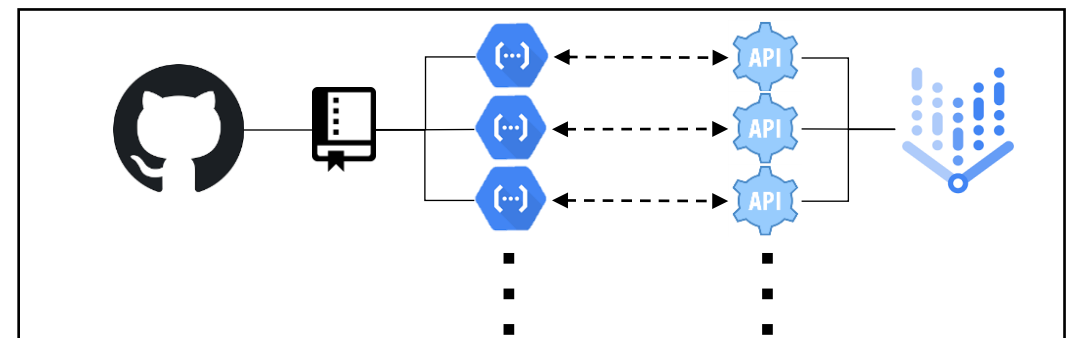
1. GCF requires much **less management** and upkeep
2. Establishes a standard method of preprocessing data for ML models → **more efficient** and *scalable* delivery of ML insights



### Previous services

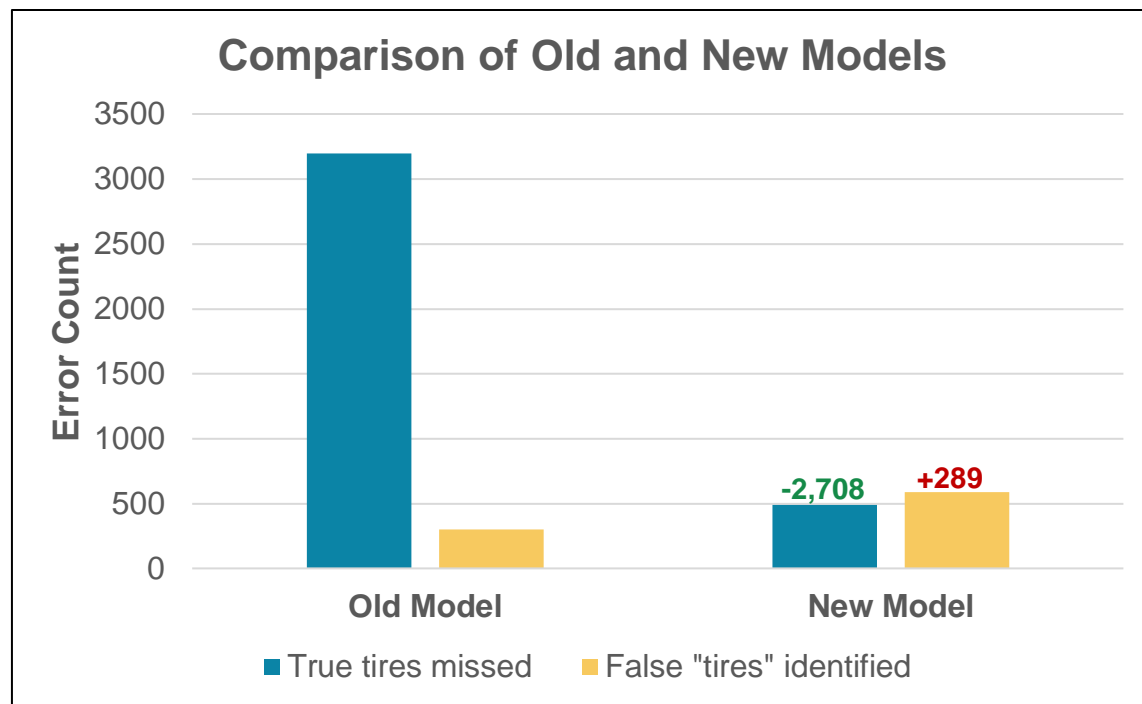


### Now with GCFs



### 3. More Accurate and Faster Insights from `is\_product`

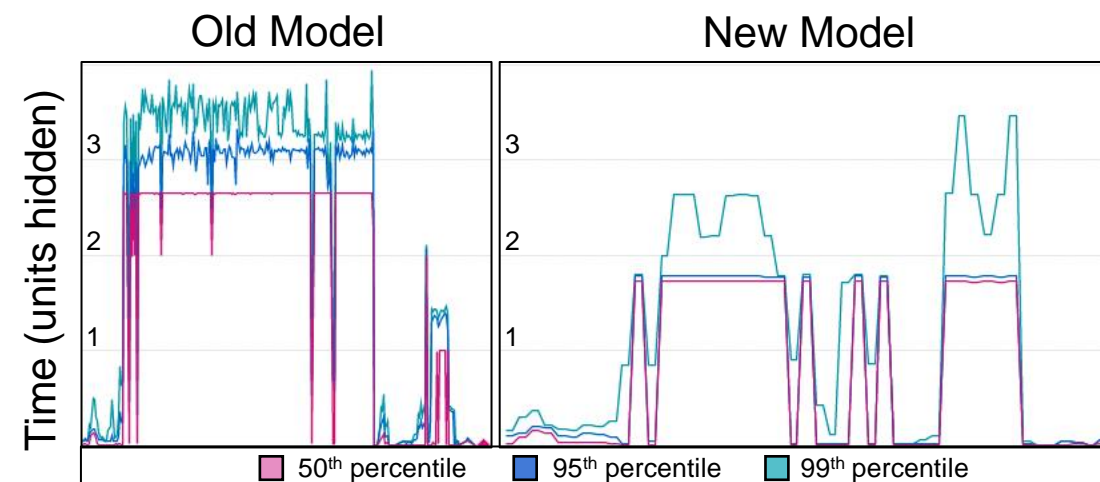
#### Evaluation and Improvement



❖ More products identified → more matches → **customer insights are more accurate**

#### Increasing Efficiency

- Online predictions are **30% faster**
- Training the model is **95% faster**



❖ More efficient execution → **lower cost**



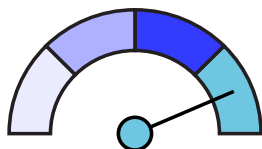
## 4. Scalable Processes Support Future Growth

### Start-to-finish Vertex pipeline in development

- Key features include...



Automatic Execution



Performance Monitoring

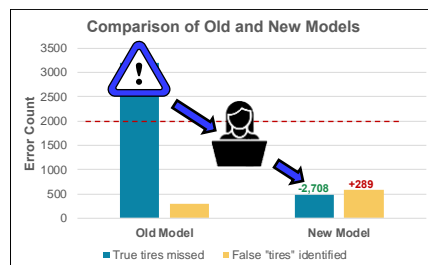


Alerts

- Allowing the company to...



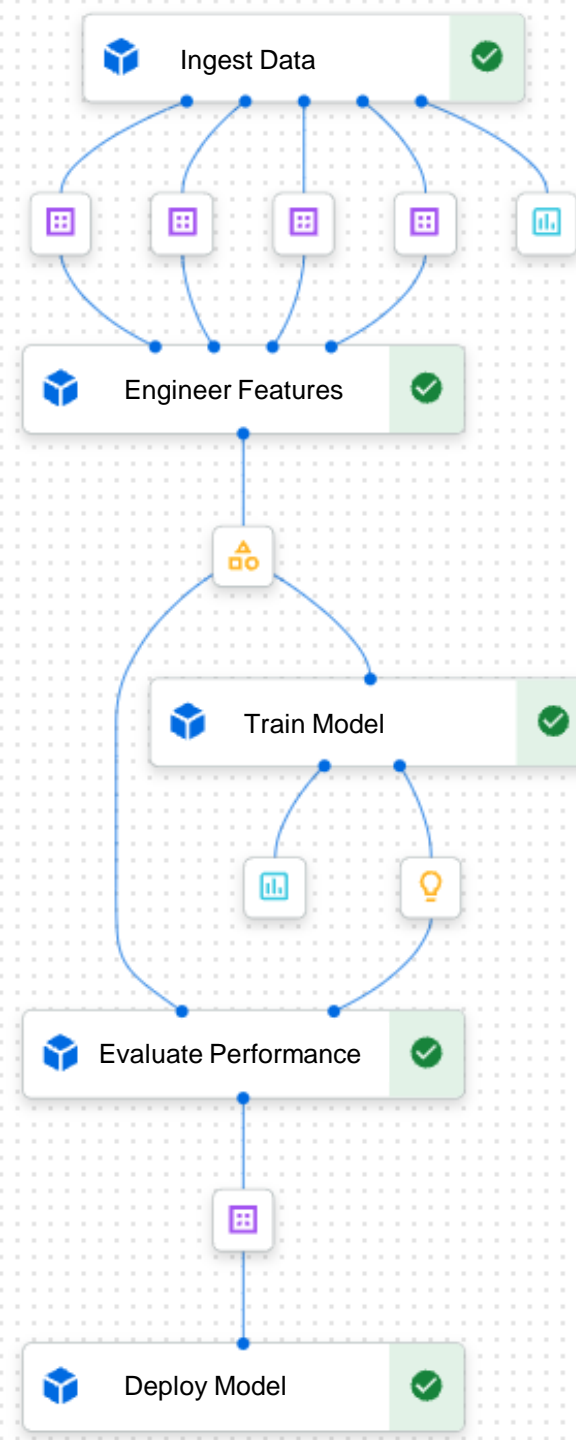
Provide more insights



Maintain high quality



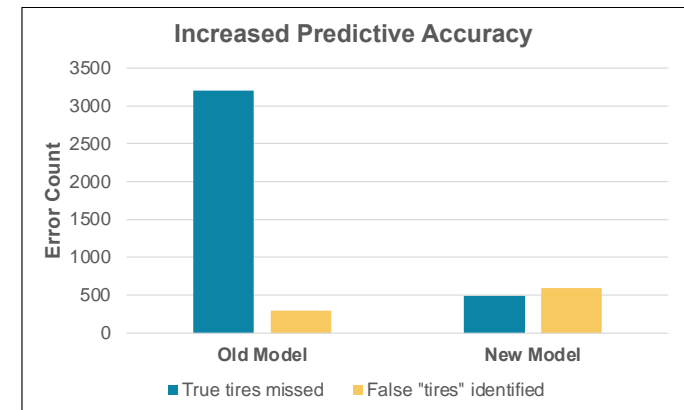
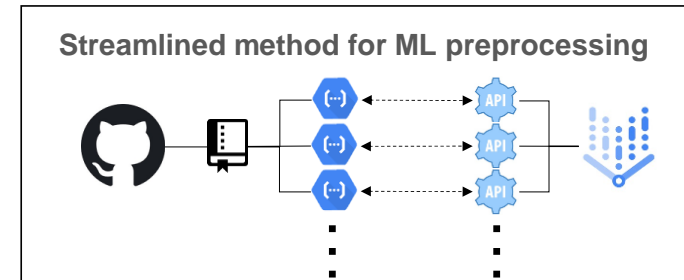
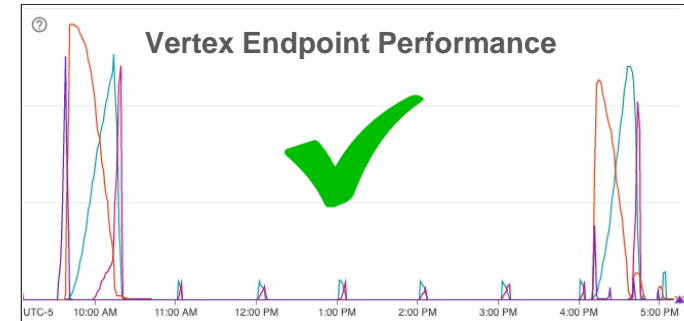
Innovate faster



# Company A is on the Right Track in Preparing for Growth

## Business Value Summary

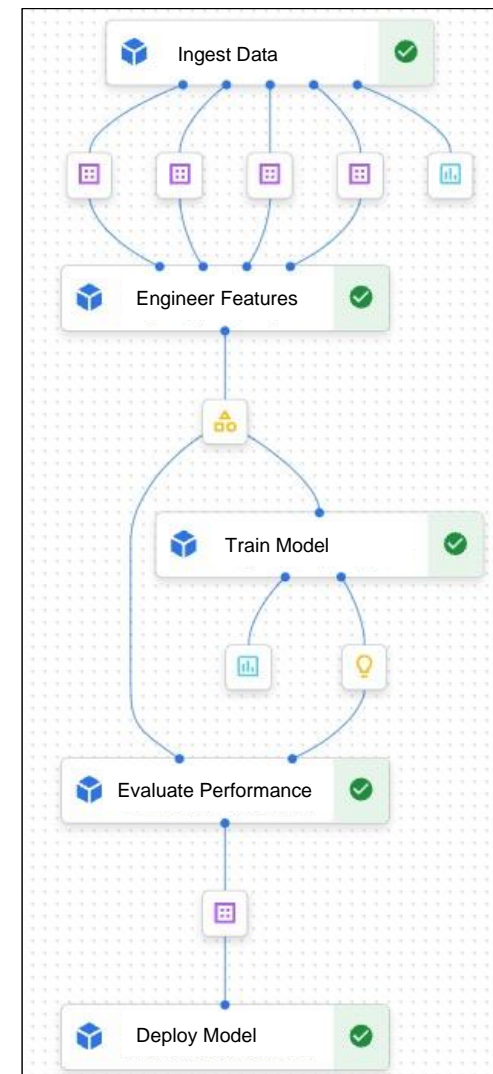
1. Confirmed **viability** of the way forward to scale ML value delivery with growing demand
2. Streamlined preprocessing—**reduced maintenance burden** and prepared to scale
3. Improved performance, **reduced cost**
4. Increased **accuracy** of product-related metrics and customers' business insights



# Continue Streamlining MLOps and Apply to New Offerings

## Future Work

1. Complete and implement the full pipeline in Vertex (in progress)
2. Apply the same structures to near-product transactions



# References

---

<https://Company A.com/>

<https://cloud.google.com/vertex-ai>

<https://databricks.com/glossary/mlops>



An aerial photograph of a two-lane asphalt road that curves through a dense, green forest. Two cars are visible on the road: one is further along the curve on the right side, and the other is closer to the text on the left side. The text is overlaid on the left side of the image.

# Questions?

Thank you for your time.