

Storage Management Software

Milestone 04

Luis Bodenbach (3201837)
Thomas Scherer (3698925)
Berkay Baltaci (3174223)
Bilal El Kadri (3854428)
Amartya C. Dudhe (3928343)

1 Introduction

Lahn Inc. runs an online shop but doesn't have much space to keep products in the storage. That means every choice about what to order and when becomes important. To solve that issue, we were provided with a dataset of previous transactions, with over 110,000 sales between 2011 and 2016. Each transaction in this file contains information like the sale date, who bought the product (age, gender), where they live, what and how much they bought, and how much money was involved. Looking at this data, some patterns can directly be seen. Sales hit their highest points in May, June, and December. Sales didn't just stay steady; they grew a little every year. This makes predicting future demand necessary if Lahn Inc. wants to avoid empty shelves or a lot of unsold products. Three major product categories make up most of their business, with 18 subcategories underneath and more than 130 products overall. Some products make much more money than others. For example, "Bikes" bring in almost two-thirds of all profit, but "Clothing" makes up just a small slice. Most buyers are in the USA, especially California, but the company also ships to five other countries, which means needed stocking might change in each country. The problem is simple: Too much stock costs money and takes up space. Too little stock means lost sales. Both are bad for the business. Our project's goal is to predict what will sell and when by forecasting demand, so storage gets used better and customers can always find what they want.

Then we want to use AI to create a Business Report with recommendations, so the company can easily get insights in only a few minutes. With smarter stocking based on the patterns we see in the data, Lahn Inc. can cut down costs and boost profits, making the whole business more profitable, run smoother and satisfy customer needs.

2 Technical Solution

Our solution was a combination of machine learning models, forecasting algorithms, and software development techniques to meet Lahn Inc.'s needs. We followed the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining) to structure the project phases. This provided a clear roadmap from business understanding and data understanding through data preparation, modeling, evaluation, and deployment.¹ Data Processing & Features: We did not perform that much data cleaning, we found around 1,000 duplicate records, but decided not to delete them, because these transactions did not have a Unique identifier, so the duplicates could have just come from different Persons with the same name, in the same country buying the same products and quantity. For predictive modeling, we implemented gradient boosting machine learning models using XGBoost. This is a high-performance libraries that can be used for regression and classification tasks. In our case, the primary prediction target was Order_Quantity (the number of units sold per transaction or aggregate period), treating it as a demand indicator. We trained the model on historical data to predict future demand

¹ Chapman et al., 2000

⁶⁷ levels. XGBoost was chosen for its ability to ¹¹⁶ make the output of our analyses more accessible
⁶⁸ handle nonlinear relationships, being very quick ¹¹⁷ we integrated a Large Language Model (LLM) to
⁶⁹ and feature interactions effectively, while also ¹¹⁸ generate polished reports. Specifically, we used
⁷⁰ providing useful outputs like feature importance ¹¹⁹ Ollama which is a runtime environment for
⁷¹ scores for interpretability. The result was a ¹²⁰ deploying LLMs locally that uses a model such as
⁷² predictive model that can forecast short-term ¹²¹ Phi-3 or LLaMA 3.2³ and even the newest
⁷³ demand for products, helping inform how much of ¹²² Deepseek R1 (which goes up to 671B and is around
⁷⁴ each item to stock in upcoming periods. However, ¹²³ 400GB)⁴. The pipeline works as follows: after
⁷⁵ XGBoost did not deliver high quality outputs, ¹²⁴ performing, for example a demand forecast for
⁷⁶ which sometimes were really questionably, so we ¹²⁵ Germany next June, the system passes the
⁷⁷ decided to use Prophet for time Series Forecasting.² ¹²⁶ forecasted quantity for each sub-category as a table
⁷⁸ Prophet is a robust forecasting library that models ¹²⁷ to the LLM as our variable part. The prompt also
⁷⁹ trend, seasonality, and holidays, and is well-suited ¹²⁸ includes some fixed, guided commands on how the
⁸⁰ for business time series data. The runtime of the ¹²⁹ pdf file should be structured and which
⁸¹ prophet model is between 1 and 2 minutes, instead ¹³⁰ components have to occur in the Report (for
⁸² of 10 seconds for the XGBoost approach. This is ¹³¹ example a business tip for the worst 3 performing
⁸³ because for instance, Prophet quantified the yearly ¹³² categories and the instruction, to create a
⁸⁴ seasonal effect (peaks around mid-year and ¹³³ professional report for a business audience) to the
⁸⁵ December) and could extrapolate trend growth. ¹³⁴ LLM. The LLM then produces a well-written
⁸⁶ The forecasts outputs of Prophet were reliable ¹³⁵ summary. By including an LLM, we essentially
⁸⁷ looking predictions which we were very happy ¹³⁶ added an AI “analyst” to our system that can take
⁸⁸ with. By predicting a few months ahead, Lahn Inc. ¹³⁷ the quantitative findings and express them in
⁸⁹ can time its reorders to ensure stock arrives before ¹³⁸ natural language, tailored to support business
⁹⁰ demand surges and avoid excess ordering when a ¹³⁹ decision-making. This automated report generation
⁹¹ dip is expected. Prophet’s ability to incorporate ¹⁴⁰ was implemented as a final step in our FastAPI
⁹² seasonality made it ideal for capturing the patterns ¹⁴¹ pipeline. Once analysis is done, the backend calls
⁹³ identified. We validated the forecasts by manually ¹⁴² the local LLM with the prepared prompt, then
⁹⁴ creating a Power Pivot table in excel, extracting the ¹⁴³ returns the generated report (PDF) to the user.

⁹⁵ information we needed to compare with the
⁹⁶ forecasts. System Architecture: The backend was
⁹⁷ built with FastAPI, a modern Python web
⁹⁸ framework, which served our trained models and
⁹⁹ forecasting logic as RESTful endpoints. This
¹⁰⁰ backend could ingest new data or parameters and
¹⁰¹ return predictions, forecasts, and recommended
¹⁰² actions, but we only used it to transfer two
¹⁰³ variables *when* and *where* from frontend to
¹⁰⁴ backend. On top of this, we created a simple front-
¹⁰⁵ end interface in two forms: a lightweight web UI
¹⁰⁶ using HTML/CSS (served via the FastAPI
¹⁰⁷ backend) for easy access through a browser, and a
¹⁰⁸ desktop UI built with Tkinter (very minimalistic
¹⁰⁹ and optically not very appealing) for demonstration
¹¹⁰ purposes. The user, like an inventory manager, can
¹¹¹ input the country and the month, that he wants
¹¹² forecasts about. The backend function is then
¹¹³ called with those parameters, and results are
¹¹⁴ directly moved into an output folder as a pdf file.
¹¹⁵ Automated Reporting with LLM Integration: To

² Taylor & Letham, 2018

³ Vaswani et al. (2017)

⁴ Online-Website: <https://ollama.com/library> (last visited: 15/07/25 at 09:25 am)

¹⁴⁴ 3 Insights from the Data

¹⁴⁵ Looking at our sales data, we immediately see that
¹⁴⁶ certain products always appear to be in the top 3
¹⁴⁷ performing sub-category. For example, based on
¹⁴⁸ several test project reports generated with our
¹⁴⁹ application, we see a high performance of “Tires
¹⁵⁰ and Tubes”, which is constantly at the top.
¹⁵¹ Showing strong and steady demand.

¹⁵² When we use Prophet for demand forecasting, our
¹⁵³ ForecastGenerator class always provides the same
¹⁵⁴ output, which looks like the following table:

Sub_Category	Forecast
Tires and Tubes	1624
Bottles and Cages	1103
Helmets	917
Caps	502
Jerseys	433
Fenders	187
Cleaners	111
Gloves	100
Hydration Packs	81
Vests	78
Touring Bikes	63
Mountain Bikes	55
Road Bikes	51
Socks	21
Bike Stands	2
Bike Racks	1

¹⁵⁵ The table ranks each sub-categories demand by
¹⁵⁶ forecasted demanded unit. This table goes directly,
¹⁵⁷ how it is, into our LLM prompt, together with some
¹⁵⁸ instruction for output generation for the LLM.

¹⁵⁹ When it comes to managing and optimizing
¹⁶⁰ inventory storage, it only makes sense that we
¹⁶¹ prioritize good performing categories. However: It is
¹⁶² obvious, that more tires and tubes are getting sold,
¹⁶³ than actual bikes. Still, 60 sold bikes make more
¹⁶⁴ profit for the company than 2000 tires, which
¹⁶⁵ means prioritization based only on demand quantity
¹⁶⁶ is nonsense.

¹⁶⁷ Back to the data, we can see that there is a pattern
¹⁶⁸ that repeats every year: sales ticking upward as
¹⁶⁹ spring turns to summer, with May and June
¹⁷⁰ especially busy. People seem to be getting ready for
¹⁷¹ warmer weather. Then, December brings another
¹⁷² rush, probably tied to holiday shopping. Sales slow
¹⁷³ down in July, August, January, and February. So,

¹⁷⁴ the business has its busy stretches in early summer
¹⁷⁵ and again at the year's end, then quiet spells right
¹⁷⁶ in the middle of summer and just after New Year's.
¹⁷⁷ When we ran demand forecasts using Prophet,
¹⁷⁸ those trends became even clearer. The model
¹⁷⁹ picked up on the spikes in sales around mid-year
¹⁸⁰ and December and the slower months in between.
¹⁸¹ Just looking at average sales wouldn't show these
¹⁸² swings. If Lahn Inc. planned only for “typical”
¹⁸³ months, they might run out of stock during peak
¹⁸⁴ times or end up with too much inventory sitting
¹⁸⁵ around when things slow down. This is why data
¹⁸⁶ analysis is very important.

¹⁸⁷ There's a sharp split when you look at profits by
¹⁸⁸ category. Bicycles make up about 64% of profits,
¹⁸⁹ accessories about 27%, and clothing just 9%. These
¹⁹⁰ numbers come from profit data collected between
¹⁹¹ 2011 and 2016. So, bicycles are the big
¹⁹² moneymakers. Clothing, while it's one of the main
¹⁹³ categories, only brings in a little money.

¹⁹⁴ Sales by region also paint an interesting picture.
¹⁹⁵ The United States leads the way, especially
¹⁹⁶ California. This is not a big surprise because it is a
¹⁹⁷ big state and the company has a strong presence in
¹⁹⁸ America. But other regions, including parts of
¹⁹⁹ Europe, have their own sales patterns. Local
²⁰⁰ holidays and weather seem to play a role. In
²⁰¹ Australia and United States, demand for bikes is
²⁰² especially high and makes up a large slice of all
²⁰³ bicycle sales (10.000 – 11.000 bike sales compared
²⁰⁴ to 2.000 – 4.000 in the other countries). So, using
²⁰⁵ the same stocking plan everywhere wouldn't work.
²⁰⁶ Different countries need different strategies.

²⁰⁷ Here for a broad overview a table on the sum of
²⁰⁸ total sold units (every product) for each country,
²⁰⁹ making it clear where the main markets and
²¹⁰ therefore the main focus of Lahn Inc. should be:

Country	Sum of sold items
Australia	263.585
Canada	192.259
France	128.995
Germany	125.720
United Kingdom	157.218
United States	477.539

²¹¹ To sum up, forecasting demand with Prophet
²¹² helped Lahn Inc. see not just when sales go up and
²¹³ down, but how that varies by category and location.

²¹⁴ These details can help the company make smarter
²¹⁵ inventory choices and stay ready for what
²¹⁶ customers want, when they want it.

218 4 Business Recommendations

219 The above insights guided us to propose the
220 following business recommendations for Lahn Inc.
221 to improve their inventory management and overall
222 profitability:

223 Predict and Prepare for Seasonal Peaks: Prioritize
224 stock availability in high-demand months. The
225 company must increase its inventory of popular
226 products (especially bikes and related accessories)
227 before May, June and December of each year. This
228 may involve increasing orders from suppliers in the
229 preceding months and making sure there's enough
230 space in the warehouse for all new products. In
231 addition, aligning the marketing campaigns and
232 promotions with these peak periods can boost sales
233 when customers are most interested. On the other
234 hand, in slower months (like January and February,
235 or late summer), Lahn Inc. can reduce inventory
236 levels to avoid overstock.

237 Focus on High-Profit Categories, Rationalize Low-
238 Profit Stock: The data clearly shows that there are
239 differences between the contributions of categories
240 to profit. We recommend shifting inventory
241 investment toward the Bikes category and top-
242 performing sub-categories. These are the ones that
243 bring in the most profits. Ensure that new models
244 or variants of bikes, which historically sell well, are
245 always stocked to our application's forecast. This
246 reduces the chance that they need to restock, which
247 can be really costly. On the other hand, they should
248 re-evaluate the Clothing category: with only ~9%
249 of profit coming from clothing, Lahn Inc. should
250 consider reducing the range or depth of inventory
251 in this category. However, not having stocked a
252 cheap product can also backfire into customer
253 dissatisfaction and effect the sales of other
254 products, so they need to be careful here.

255 Implement Regional Inventory Strategies: adapt
256 inventory levels and assortment to regional
257 demand patterns. Given that the USA (especially
258 California) is a sales driver, warehouses serving
259 that region should maintain higher inventory levels
260 for the top-selling products (such as the most
261 popular bike models) and prepare for the
262 significant increase in demand during the summer
263 and holidays in that region. In regions where
264 demand is lower or where seasonal patterns differ,
265 adjust the stock accordingly. This could involve a
266 slightly later build-up for holiday inventory in
267 Europe if their peak is more concentrated in

268 December than November.

269 Adopt Demand-Driven Reordering (Automation):

270 We suggest adding our forecasting system to the
271 reordering process. Instead of always ordering
272 manually, Lahn Inc. should use the results from our
273 model to trigger reorders. For example an
274 automated alert system: if the forecast for the next
275 N weeks shows that the stock of Product X will fall
276 below a critical level before the next shipment can
277 arrive, an order is suggested or automatically
278 placed with the supplier. This kind of just-in-time
279 inventory management will keep inventory levels
280 just right, never too high which costs money and
281 never so low that it causes stockouts. It may also be
282 helpful to shorten supplier lead times or find other
283 suppliers for important items (especially those in
284 the Bikes category). This will help respond to
285 sudden increases in demand. Over time, this
286 demand-driven approach can be improved to
287 include safety stock calculations for products with
288 forecast uncertainty. The goal is a semi-automated
289 system where the data triggers inventory actions.
290 This frees up the managers to handle exceptions
291 and strategy instead of manual monitoring. Should
292 Lahn Inc. decide to move forward with this project,
293 we would be more than ready to take on the
294 assignment.

295 Bundle for complementary Products: use product
296 pairings to create bundled offerings and improve
297 the customer experience. When the data shows that
298 people often buy a bike and a helmet together, Lahn
299 Inc. can formalize that by selling the "Bike and
300 Helmet Combo" at a small discount, or by just
301 suggesting the helmet on the bike's product page.
302 Bundling not only can increase sales of accessories
303 but also ensures customers get everything they
304 need at once, improving satisfaction.

305 5 Reflection on Plan vs. Implementation

306 At the beginning of the project, our goal was to
307 build a system that could help improve inventory
308 decisions using data. We planned to use machine
309 learning models like XGBoost or LightGBM to
310 predict demand and evaluate product profitability.
311 The idea was also to build a user interface where
312 people could select different types of analyses and
313 get readable summaries, generated with the help of
314 large language models such as LLaMA or Phi-3.
315 We wanted the system to do more than just
316 calculate numbers and it should explain them in a
317 clear and helpful way.

318 As we got further into development, we adjusted
319 our approach in a few areas. Instead of using

320 XGBoost or LightGBM for the demand forecasting 372 In conclusion, the project ended up in a different
321 part, we switched to Prophet. This decision was 373 place than we first expected, but it still achieves the
322 based mostly on practicality. Prophet works 374 original goal: helping businesses plan better by
323 especially well for time series forecasting, like 375 turning raw data into useful information. We
324 predicting future sales by month. It's easier to use 376 learned that it's important to be flexible, especially
325 than the tree-based models we initially planned, 377 when working with real data and limited time. The
326 and it handles seasonal patterns and trends 378 tools we used changed, and some features were
327 automatically. That made it a good fit for the kind 379 made simpler or delayed, but we always kept the
328 of grouped data we were working with. 380 main goal in mind. The end result is a working

329 A notable advancement in comparison with our 381 system that makes forecasts, explains them clearly,
330 initial proposal is the implementation of two 382 and provides a good starting point for further
331 distinct user interface options: a desktop graphical 383 development.

332 user interface and a web-based interface. Both

333 versions allows the user to select a country and a

334 month, after which the system processes the data,
335 generates a demand forecast, employs a language

336 model to create a structured business report, and
337 finally saves the result as a PDF. This dual interface

338 approach offers a high degree of flexibility in terms
339 of the system's application, depending on user

340 preference or technical configuration.

341 Part of the original plan that was implemented was
342 the use of a large language model. We connected

343 Ollama to the system and used the LLaMA 3.2
344 model to convert forecast data into a clear,

345 structured report. The generated text is included in
346 the final PDF file, making the results easier to

347 understand for non-technical users. This part of the
348 project worked well and largely met our original

349 expectations.

350 We ultimately decided not to implement a
351 profitability classifier or market basket analysis,

352 even though these were part of our original
353 concept. Given the time constraints and the

354 complexity involved, we chose to focus our efforts
355 on demand forecasting and report generation—

356 features that were more central to the project's main
357 goal. Financial attributes such as profit and price

358 were also left out in this phase, as they were more
359 relevant to profitability assessments than to the

360 kind of demand predictions we were prioritizing.

361 In the end, these choices proved to be reasonable.

362 With Prophet, it enabled us to quickly generate
363 meaningful forecasts, gives us more time to focus

364 on integrating the language model and building a
365 reliable reporting system. Although the final

366 version does not include all the features we initially
367 planned, it provides a clear and functional

368 prototype of the system. The core idea has been
369 fully implemented and is working well, and the

370 features we left out can still be added in future
371 iterations without the need for a complete redesign.

384 References

385 Chapman, P., Clinton, J., Kerber, R., Khabaza, T.,
386 Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-
387 DM 1.0: Step-by-step data mining guide. SPSS Inc.

388 Taylor, S. J., & Letham, B. (2018). Forecasting at
389 Scale. *The American Statistician*, 72(1), 37-45.

390 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J.,
391 Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017).
392 Attention is all you need. *Advances in Neural
393 Information Processing Systems*, 30, 5998-6008.

394 Online-Website: <https://ollama.com/library>