Slide 1 Ivan

Good afternoon, everyone!
Thank you for being here. My name is Ivan Chertov, and this is my colleague, Emma Le Bars.

Today, we're excited to present our project: **Demand Forecasting and Inventory Management for H&M**.

We've taken a data-driven approach to optimize warehouse space and stocking.

Let's dive in!

Slide 2

Before we dive into the details, let me guide you through the structure of our presentation.

We will begin with an overview of the **business context** and the objectives of our capstone project.
Next, we'll explore the **data** we worked with and the insights we derived during the exploratory analysis.

We will then get to the flesh of our project:
the **prediction models** we used and of course
share our **results** with you.

After that, we'll introduce our **inventory rebalancing strategy.**

Finally, we'll wrap up with the **key insights** and takeaways from our project.

Let's get started with the business context!

Slide 3

To set the stage, let's talk about the **business context** and objectives of our project.

Our primary **objective** is to provide accurate demand forecasts for H&M. This will help optimize inventory space and ensure better stock management.

Now, why is this important?
Effective stock management prevents overstocking and shortages. By aligning inventory levels with actual customer demand, H&M can improve sales performance while simultaneously reducing storage costs.

Our **primary goal** for this project is to develop a forecasting model that identifies high-demand products, especially during peak seasons. This model also supports a strategic inventory rebalancing process.

This is the foundation of our work. Next, we'll explore the data that drives this solution.

Slide 4

Let's move on to the data we used for this project.

The dataset includes **transaction data** from H&M's online sales. It spans from **September 2018 to September 2020** and is publicly available on Kaggle. We have a big dataset with over 30 million transactions and 104000 articles.

We focused on following **key attributes** of the dataset, which are:

Product type, color, weekly sales (total units sold), average price, and customer counts. Let's have a look at the data itself.

And at this point, I give the floor to my valued colleague, Emma

Slide 5

Thanks Ivan! Hello everybody!
This chart illustrates **total units sold over time**, showcasing fluctuations in monthly sales.
Key highlights include:

- A significant spike in **June 2019**, reaching nearly **1.8 million units**, potentially due to seasonal demand or promotional campaigns.
- A notable decline in **September 2020**, with sales dropping to just over **500,000 units**, likely influenced by external factors such as market shifts or reduced inventory.

Slide 6

When analyzing product sales, we found that **dresses, sweaters, and trousers** consistently dominate our overall sales.

These items account for the largest share of units sold, making them a key focus for inventory planning and demand forecasting.

Meanwhile, categories like skirts and shorts show lower sales volumes, but they may still see spikes during specific seasons.

Slide 7

And also based on colour we see that Trousers, Dresses and Sweaters dominate.

Black, Beige and Blue are the favourite colours with **black trousers** being the **best selling category**.

Slide 8

Here, we analyzed **seasonality trends** across Top 10 Best selling product categories.

As shown in the stacked bar chart, categories like **trousers, dresses, and sweaters** not only dominate overall sales but also show distinct seasonal variations.
For example, **sweaters peak in fall and winter**,
while **dresses and T-shirts sell more during spring and summer**.

This seasonality insight is crucial for inventory planning.
By understanding these patterns, we can better align stock levels with seasonal demand, reducing overstocking and shortages.

Next we gonna have a look at our prediction models and I give the floor back to Ivan

Slide 9

Thanks, Emma!
Now, let’s take a look at how we built our prediction model.
To achieve reliable forecasts, we turned to machine learning and dove into exploring various approaches!

Slide 10

To identify the best approach for our forecasting model, we took a look at several options:

1. **Random Forest:** While simple and interpretable, its predictions were not accurate enough for our needs.
2. **XGBoost:** This boosted algorithm showed promise but was computationally intensive, straining our resources.
3. **A combined approach using Random Forest, XGBoost, and PyTorch:** Although this hybrid method offered advanced capabilities, it proved too CPU-heavy for practical application.

Through these iterations, we learned the trade-offs between performance, accuracy, and resource efficiency, guiding us toward more feasible solutions

We got into a frustration zone, but we never give up!

Slide 11 + Slide 12

And finally we created our Ensemble Model!

Containing of Random Forest Regressor,
Linear Regression and MLP Regressor!

The Meta model of our Ensemble is Random Forest
What a meta model does is basically
combining predictions from base models,
it learns to weight them based on performance,
and delivers more accurate final predictions.
And now Emma will tell you more about training of the model!

Slide 13

For our model, we split the data into four parts to ensure robust training and accurate testing:

1. **Training Set**: Data before January 2020 was used to train the model on historical patterns
2. **Validation Set**: January 2020 data allowed us to fine-tune the model's parameters and check its performance
3. **Test Set**: February 2020 data was used to evaluate the model's accuracy and reliability on unseen data
4. **Forecasting Set**: For March 2020, we focused on generating forecasts while excluding data influenced by the COVID-19 pandemic to maintain accuracy

Slide 14

In February 2020, our model achieved the following results:

- It predicted **881,000 units** of sales, compared to the actual **875,000 units**, resulting in a small **positive variance of 0.68%**

These results show the model’s strong accuracy and reliability in forecasting demand

- The average size of the errors was **24 units**, and the overall error was **76 units**

Slide 15

This chart shows the heart of our capstone project!
This is the **forecast results of our model** compared to actual sales across various product categories.

The beige bars represent **actual units sold**,
while the red dots indicate our **predicted values**.

As you can see, the predictions align closely with actual sales for most categories, particularly in high-demand items like **trousers, sweaters, and dresses**.
As you can see our model over predicted the sales on Sweaters, but we think that the demand went down due to lockdowns in winter 2020.

This demonstrates the model's effectiveness in capturing demand trends and its potential to support accurate inventory planning.

And now Ivan will walk you through the intentory rebalancing strategy

Slide 16

Thank you Emma! So! Let’s elaborate on our **inventory rebalancing strategy** based on the results of our model.

It has two main objectives:

1. **Freeing up warehouse space** by removing low-demand items.
2. **Prioritizing high-demand products** to ensure better stock availability and efficient resource allocation.

This strategy ensures that the warehouse operates efficiently, aligning inventory levels with customer demand to reduce both overstock and shortages.

Slide 17

Our **Rebalancing Mechanism** is built with three thresholds:

1. A Threshold for Non-Peak Items
2. A Threshold for Peak Items
3. And a Dynamic Threshold

We implemented these by

1. Appling a **stricter threshold** of 50% of average demand to to low-demand products. Enabling quicker rebalancing and freeing up space.
2. A **threshold** of 100% is used for high-demand products to ensure we maintain sufficient stock for top-performing items.
3. And we calculate **rolling demand over a 4-week window** to adjust stock in near-real time, allowing dynamic responses to changing demand patterns

Slide 18

To monitor the effectiveness of our rebalancing strategy, we created this Rebalance Dashboard.

And let me show you how it helps in inventory management:

- Like for **Swimwear Tops**, the dashboard shows no stock increase is needed for the first three weeks, but sales are expected to rise in the final week. This allows proactive restocking
- For **Moccasins**, they are flagged across all four weeks, indicating consistently low performance. This signals a need for re-evaluation of their stock levels or potential removal

By identifying these trends weekly, the dashboard supports targeted decisions, ensuring space is used efficiently and demand is met effectively.

Back to Emma!

Slide 19

Thank you Ivan!

So let’s dive into the Key Insights of our capstone project

These are

- **Seasonal Demand Sensitivity**
- **Space Optimization**
- **Refinements & Future Enhancements**

Let’s dive into each of these in more detail

Slide 20

Our first key insight is the importance of **seasonal demand sensitivity**, which we recommend leveraging for better forecasting.

The **is_peak_season flag**, proved to be useful!
The model effectively captured demand fluctuations, ensuring it responded dynamically to both **peak and off-peak seasons**.

We suggest to integrate this feature into the demand planning to align inventory more closely with seasonal trends, reducing overstock during low seasons and ensuring availability during peaks.

Slide 21

One of our key insights is the importance of **space optimization**, which we strongly recommend implementing.

Using the **forecasts generated by our model**, H&M can focus on **removing low-demand items** to free up valuable warehouse space.

This space can then be utilized to **prioritize high-demand products**, ensuring better stock availability and reducing overstock issues.

By adopting this approach, H&M can align inventory practices with their **long-term stock management goals**, driving efficiency and profitability

Slide 22

Our recommendations for the next steps include:

1. **Further tuning of peak season thresholds** to enhance sensitivity to demand fluctuations. Additionally, exploring external factors like **marketing campaigns** could improve forecast accuracy.
2. **Integration of economic and promotional data** to refine predictions further, ensuring the model captures external influences on customer behavior more effectively.

But there is some Bonus Material from Ivan

Slide 23

Yes we have some Bonus material!

Slide 24 - Slide 27

Another area for improvement in the next steps is the **Model Training for Single Articles**.

Towards the end of the project, we began implementing this approach and achieved promising training results. However, the predictions didn't meet expectations. Fine-tuning the model required significantly more processing power than we had available.

This is an area where future efforts can focus on enhancing performance and accuracy.

Slide 28

We’ve reached the end of our presentation!

Thank you for your attention.
Feel free to ask any questions later during the breakout session
and also feel free to explore our work further
by visiting our GitHub repository.

Good bye!