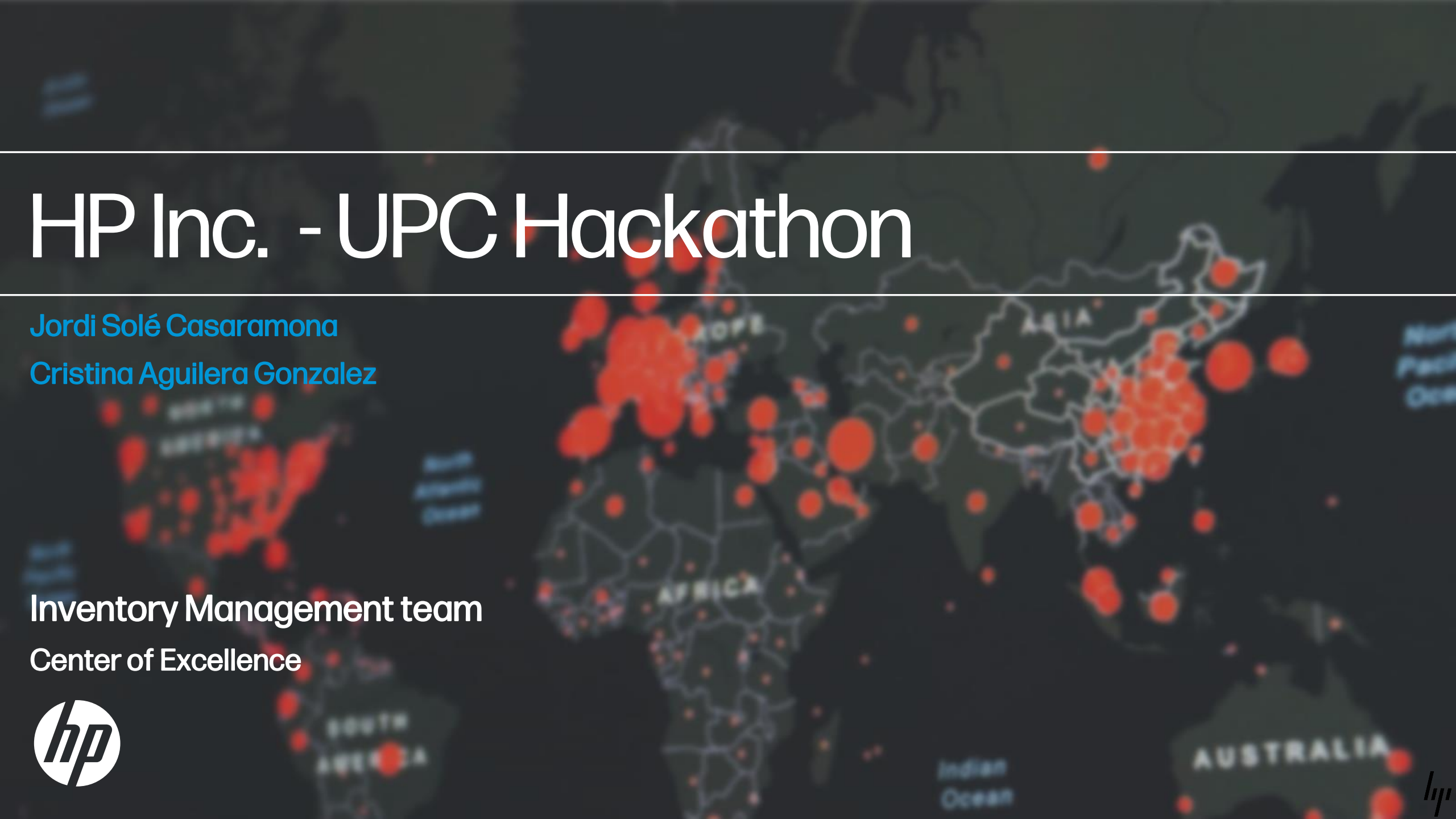
HP Inc. - UPC Hackathon

Jordi Solé Casaramona

Cristina Aguilera Gonzalez

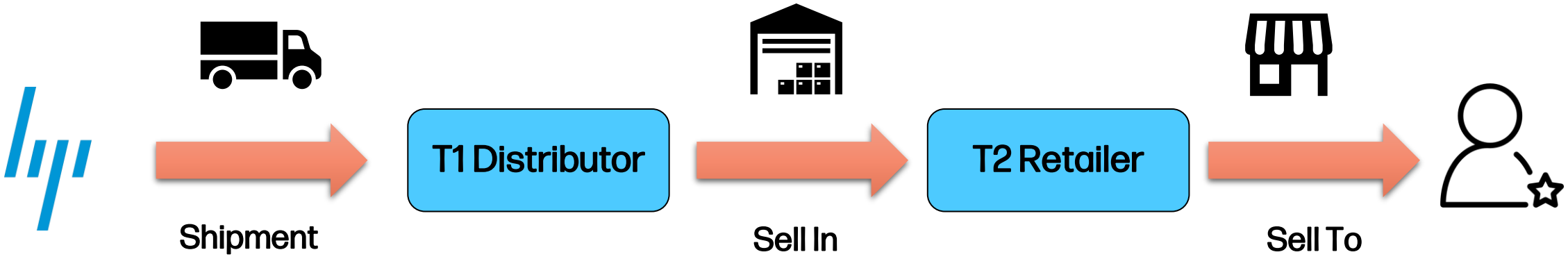
Inventory Management team

Center of Excellence



Predict Sell To, Sell Thru and Shipment

Help HP to perfect the Supply Chain Management



Why is this important?

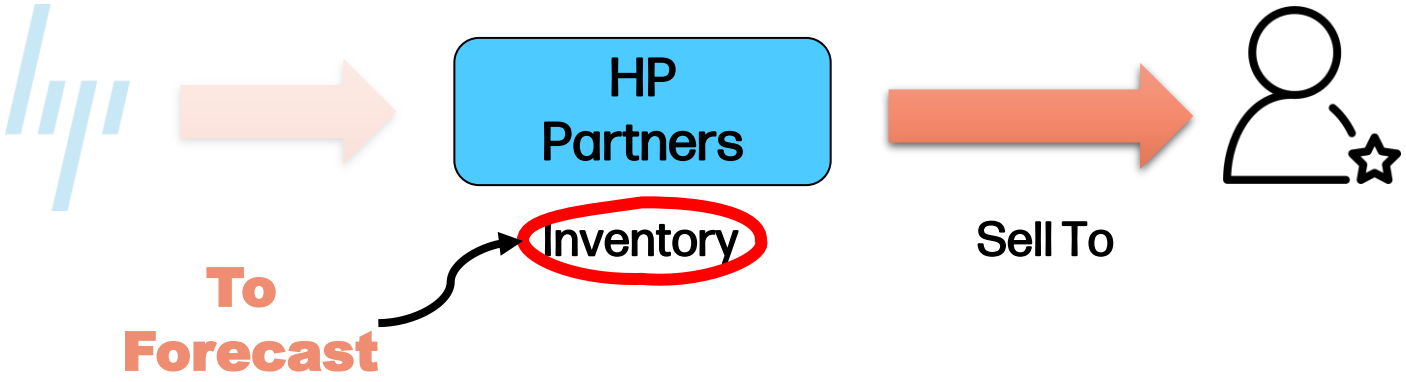
- Know how much we will sell
- Know how much we need to have
- Know how much we need to produce

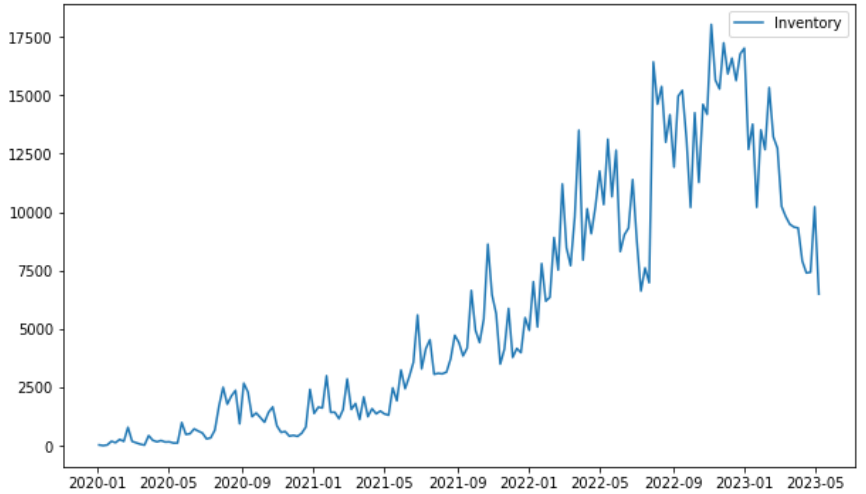
→

- Optimize Supply Chain

Hackathon Proposal

Problem subset with Anonimized and Transformed data





Why do we need you?	To Forecast Weekly Inventory for multiple Products
What do we want?	New approaches to solve the problem, free to solve it in your own way
Areas to work	Advanced Regression algorithms, clustering, Word2Vec...

What is it for you?

Great prices &
Internship/Graduate posibilities!

More information will be given at the session after this presentation.

Stayed tunned & good luck!

Friday 12th

11pm Presentation

Agenda

Who are we?

+2600
employees

+62
different nationalities

35 avg age

Largest HP R&D
Lab outside the US

+700 R&D engineers

+150 patents per year

12
different businesses

WW HQ of the 3D Printing and Large Format Printing Businesses

EMEA HQ
of the Graphic Solutions Business

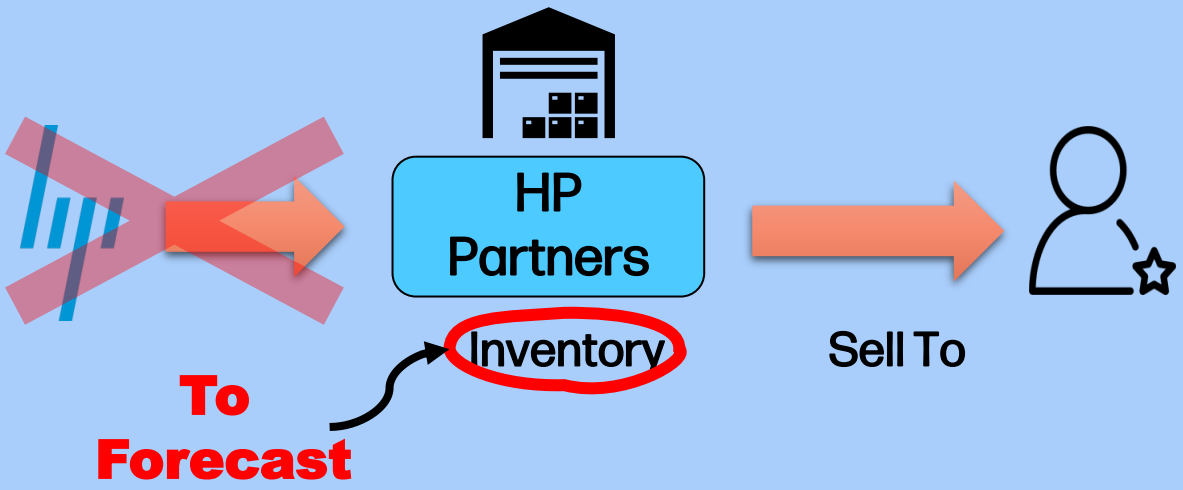
HP Team

The volunteer team that made this possible

Abel Martinez	Software Engineer
Arnau Canyadell	Software Engineer
Marco Gambarotto	Operations Manager
Luis Moro Carrera	ML Engineer
Macarena Vargas	Industrial Engineer
Laura Fernandez	Human Resources
Esteve Tarragó	Firmware Engineer
Marina Ramon	Firmware Engineer

Gabriel Andrei	Data Engineer
Cristina Aguilera	Data Science
Aleix Pujol	Data Science
Pol Monroig	Data Science
Jordina Torrents	Data Science
Tania Chubey	Data Science
Jordi Solé	Data Science
Joan Morales	Data Analyst

Business Problem



Objective

Forecast the Inventory for each of the 100 products for the next 13 weeks

+ Your Learninig!

IMPORTANT INFO

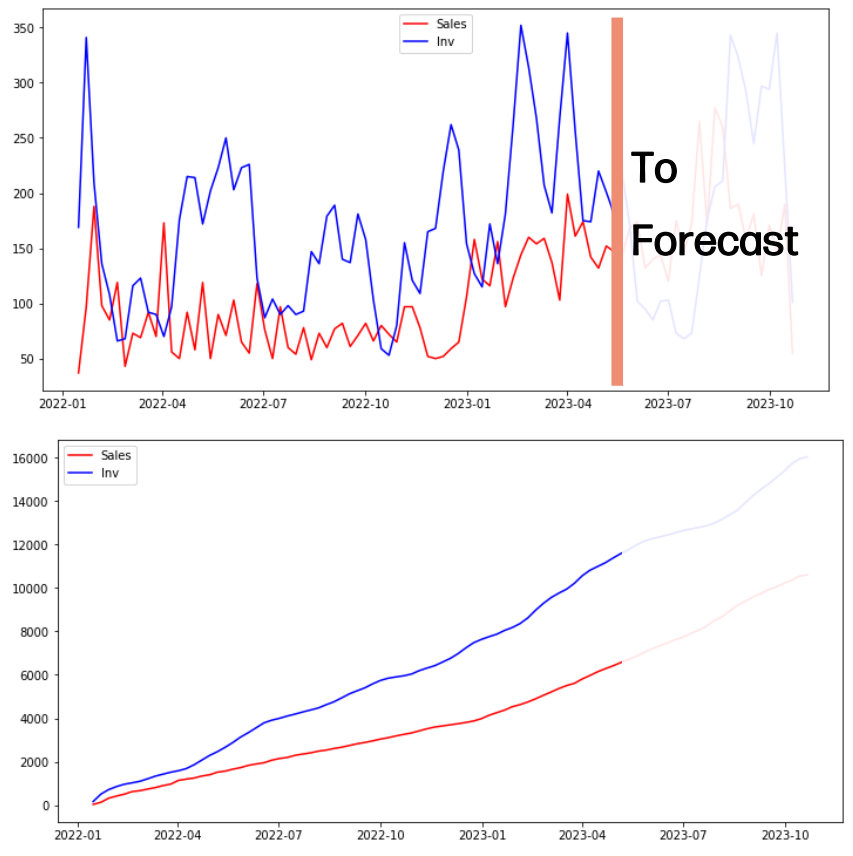
- Data is not perfect – missing values, duplicates, errors, etc.
- Sales should decrease the inventory in the same week.
- You don't have the future sales, since they have not happend.
- Cutoff is this current week. (May 7th, 2023)
- Data is ficticious.

HOW?

- You are free to do any approach you want
- No restrictions, be creative
- Metric to optimize: RMSE

Examples of the data

May 7th, 2023 is the cutoff – More information on Kaggle



Train

	id	date	year_week	product_number	reporterhq_id	prod_category	specs	display_size	segment	sales_units	inventory_units
0	202119-6909	2021-05-15	202119	6909	93	Arale	21274	13.3	Premium	2.0	35.0
1	202120-6909	2021-05-22	202120	6909	93	Arale	21274	13.3	Premium	0.0	70.0
2	202121-6909	2021-05-29	202121	6909	93	Arale	21274	13.3	Premium	3.0	137.0
3	202122-6909	2021-06-05	202122	6909	93	Arale	21274	13.3	Premium	0.0	274.0
4	202123-6909	2021-06-12	202123	6909	93	Arale	21274	13.3	Premium	0.0	333.0
...	...	...	...	...	...	...	...	...	...	...	...
15449	202317-247737	2023-04-29	202317	247737	15	Arale	10637	13.3	Premium	0.0	62.0
15450	202317-247737	2023-04-29	202317	247737	24	Arale	10637	13.3	Premium	0.0	3.0
15451	202318-247737	2023-05-06	202318	247737	3	Arale	10637	13.3	Premium	22.0	0.0
15452	202318-247737	2023-05-06	202318	247737	15	Arale	10637	13.3	Premium	68.0	78.0
15453	202318-247737	2023-05-06	202318	247737	24	Arale	10637	13.3	Premium	5.0	1.0

Test

	id	inventory_units
0	202319-105609	?
1	202319-107583	
2	202319-10857	
3	202319-109557	
4	202319-112518	
...	...	
1295	202331-92778	
1296	202331-93765	
1297	202331-94752	
1298	202331-9870	
1299	202331-99687	

Evaluation

5 minutes presentation to show your work

Methodology, Results, Visualizations & Conclusion
(WE ACCEPT SLIDES!!!)

Assessment Indicators	Indicator Questions	Weight
Solution Idea	How well does the proposed solution address the need of the challenge? Does the pipeline followed make sense in a DS framework?	15%
Implementation	Can the tool be successfully installed and used in a real business scenario? Training speed of the model. How well does the tool achieve its stated objectives? Clean code with classes	15%
Technical Innovation	How novel is the code? Is it a straightforward extension of a basic or state-of-the-art model? Are they using advanced ML tools to preprocess the data?	20%
Creativity	The solution explore creative methodologies to solve the problem vs just modelling and tunning error and goes further in understanding the data Display additional original ideas aside of the forecast, i.e., a GUI, recommendations, etc. Story telling.	25%
Errors Metrics	RMSE on testing set (Kaggle) Only prediction **Draws will be decided with this Indicator**	25%

Kaggle

Real time feedback

- The leaderboard will be public
- **Its 25% of the evaluation**
- The score is partial, not computed with the whole test set → **Prevent Overfitting**
- Thus, the final score will be know when the hackathon finishes

Metric to optimize: RMSE

SCAN ME!

Get access to the Kaggle competiton

Kaggle

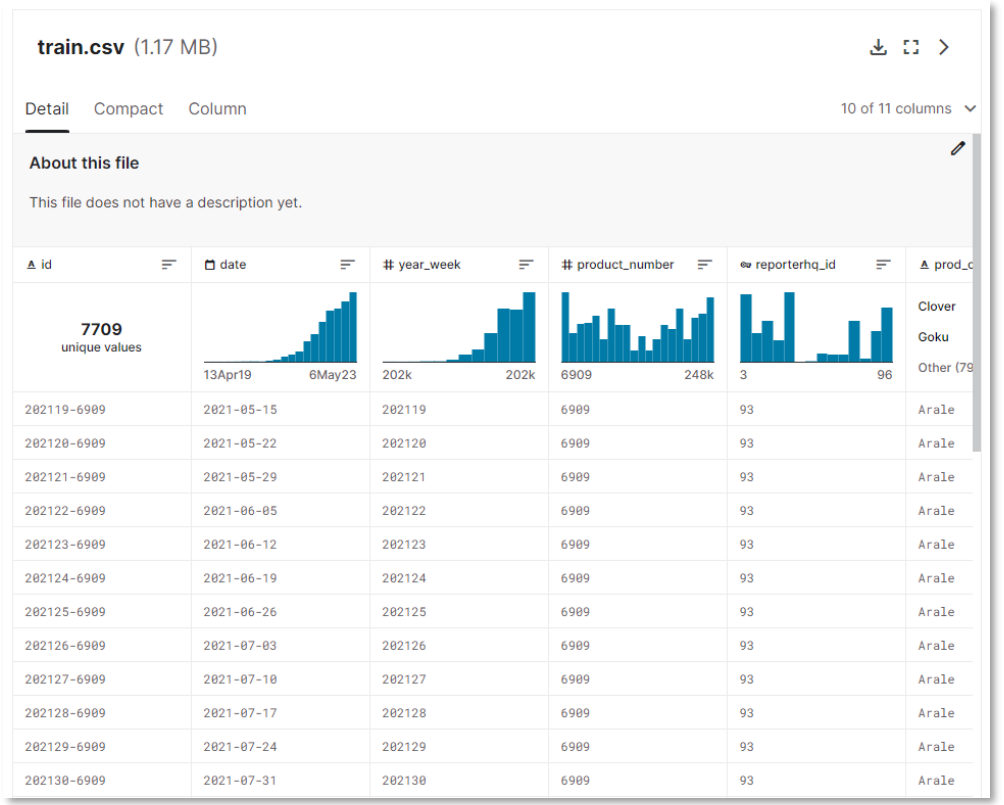
Column description

Files

- **train.csv** - the training set.
- **test.csv** - the test set.
- **sample_submission.csv** - a sample submission file in the correct format. Remember to save the csv using `, index=False`.

Columns

- `id` - `yearweek` + `-` + `product_number`. Unique identifier of the time series.
  - For example, for the first week of 2023 and the product "1234567" the `id` would be: 202301-1234567
- `date` - YYYY-mm-dd
- `yearweek` - YYYYww
- `product_number` - Unique ID for each product.
- `reporterhq_id` - Unique ID for each reseller (vendor).
- `prod_category` - Product category with a funny name that groups products from the same product line together.
- `specs` - Specifications of the product. This can be the RAM + Graphics card, or other components.
- `display_size` - Display size or screen of the PC.
- `segment` - Target segment of the product.
- `sales_units` - Sales to the final customer of the product for that week.
- `inventory_units` - Target variable. Inventory for each product_number and reporterhq_id for a specific week.



train.csv (1.17 MB)

Detail | Compact | Column | 10 of 11 columns

About this file

This file does not have a description yet.

id	date	year_week	product_number	reporterhq_id	prod_c
7709 unique values	13Apr19 – 6May23	202k – 202k	6909 – 248k	3 – 96	Clover, Goku, Other (79
202119-6909	2021-05-15	202119	6909	93	Arale
202120-6909	2021-05-22	202120	6909	93	Arale
202121-6909	2021-05-29	202121	6909	93	Arale
202122-6909	2021-06-05	202122	6909	93	Arale
202123-6909	2021-06-12	202123	6909	93	Arale
202124-6909	2021-06-19	202124	6909	93	Arale
202125-6909	2021-06-26	202125	6909	93	Arale
202126-6909	2021-07-03	202126	6909	93	Arale
202127-6909	2021-07-10	202127	6909	93	Arale
202128-6909	2021-07-17	202128	6909	93	Arale
202129-6909	2021-07-24	202129	6909	93	Arale
202130-6909	2021-07-31	202130	6909	93	Arale

Other Interesting Resources

Use this links to get familiar with time series

Store Sales TS Forecasting - A Comprehensive Guide

https://www.kaggle.com/code/ekrembayar/store-sales-ts-forecasting-a-comprehensive-guide

Introduction to Time Series Forecasting

https://towardsdatascience.com/introduction-to-time-series-forecasting-7e03c4bd83e0

Time Series Forecast : A basic introduction using Python

https://medium.com/@stallonejacob/time-series-forecast-a-basic-introduction-using-python-414fcb963000

Good luck!