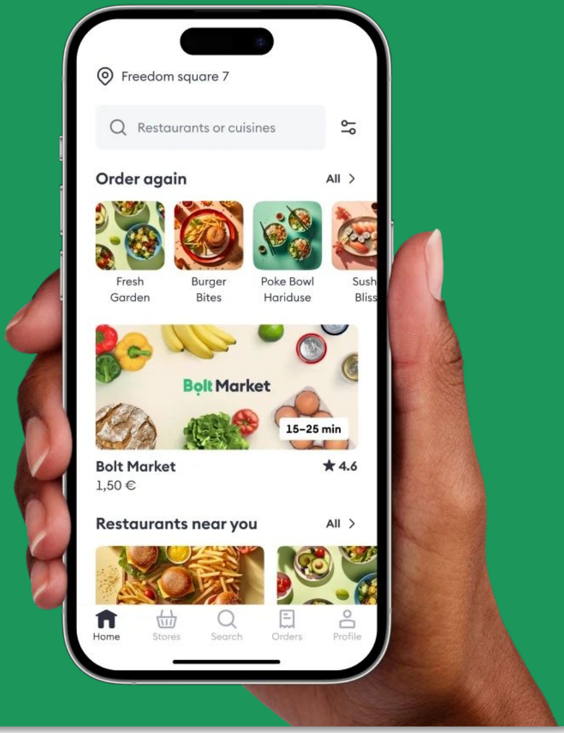
Bolt

Home task / Senior Data Analyst

Business Analytics & Commercial (Delivery)



Ivan Kartavyi / Lisbon

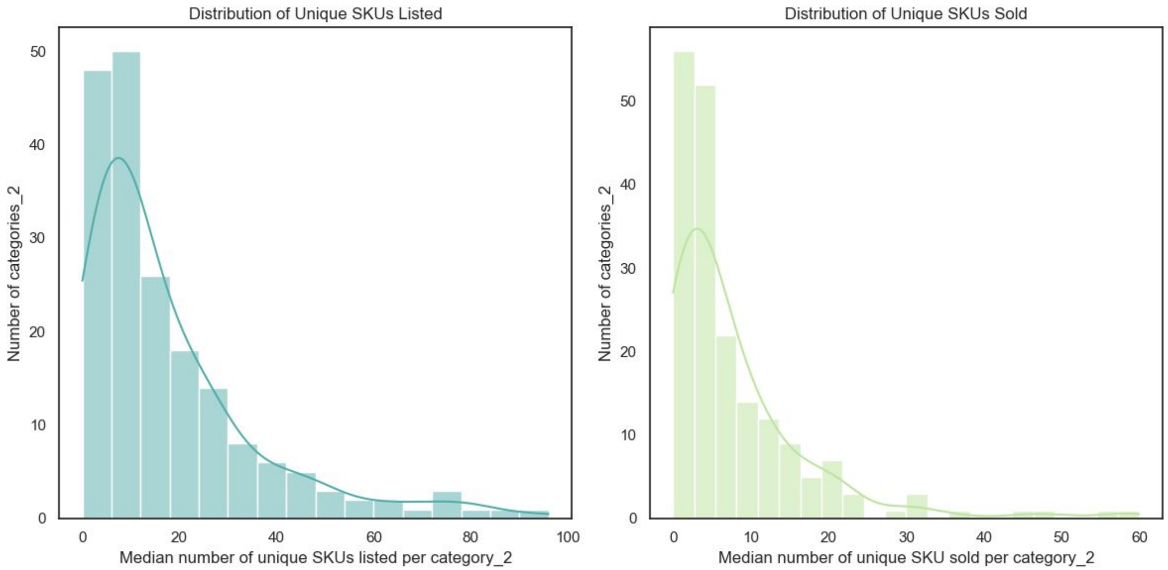
Task 1, city-level data about grocery stores SKUs

You have city-level data about 4 online grocery stores SKUs (inventory, sales, waste). Based on the information there, please provide Commercial and Supply chain teams with relevant insights you extracted from the dataset.

Also provide recommendations and hypotheses to validate further.

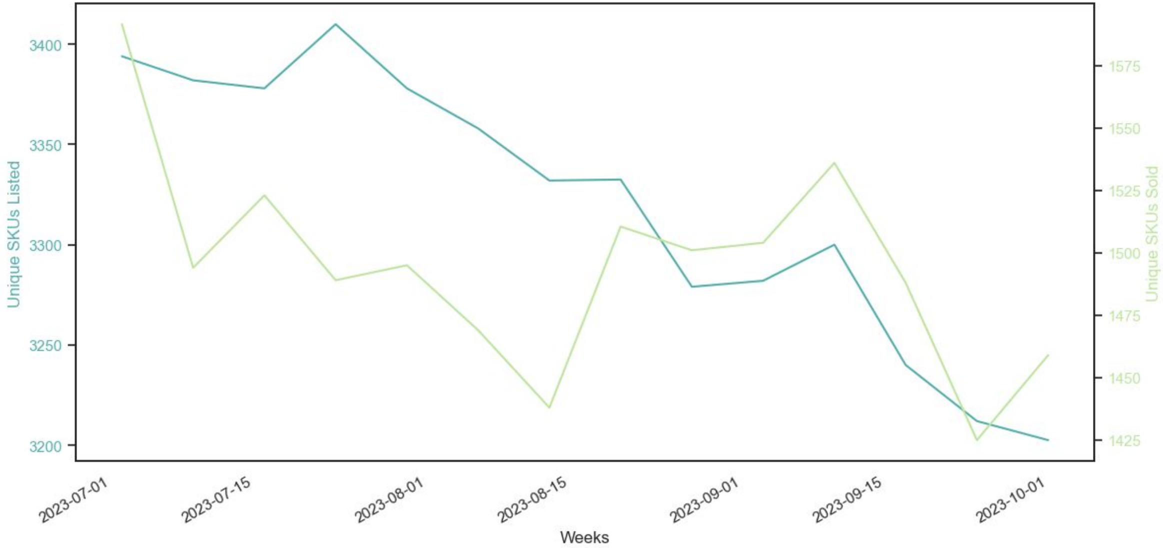
Commercial Team /
Optimizing SKU Performance in Categories

Commercial Team / Optimizing SKU Performance in Categories



Most categories aren't overloaded, so user choice isn't too dispersed across SKUs.

Commercial Team / Optimizing SKU Performance in Categories



However, the dynamics of unique SKUs sold, while positively correlated, don't always match the listing dynamics. The Pearson correlation coefficient is 0.48.

Commercial Team / Optimizing SKU Performance in Categories

This means we have categories with many SKUs but no user demand, cluttering the product list, potentially confusing users, and overly focusing their attention on top SKUs. This could increase bounce rates and impact revenue.

category_level_2	unique_skus_listed	unique_skus_sold	share
Red wine	96.0	8.0	8.0%
White wine	85.0	18.0	21.0%
Heated tobacco	80.0	30.0	38.0%
Chips	77.0	60.0	78.0%
Cigarettes cigars	77.0	57.0	74.0%
Chocolate tablets	72.0	29.0	40.0%
Toys	68.0	1.0	1.0%
Yogurt pudding	65.0	45.0	69.0%
Soft drinks	62.0	49.0	79.0%
Sauces	58.0	20.0	34.0%
Stationery	55.0	1.0	2.0%
Bacon cold cuts	51.0	36.0	71.0%

** The screenshot shows median values for each category.*

Commercial Team / Optimizing SKU Performance in Categories

Proposed solutions include:

1) Investigating these cases further and experimenting with algorithms to adjust sorting/grouping in overloaded categories.

2) Modifying the UI to organize and represent products in these categories more clearly for the user.

	unique_skus_listed	unique_skus_sold	share
category_level_2			
Red wine	96.0	8.0	8.0%
White wine	85.0	18.0	21.0%
Heated tobacco	80.0	30.0	38.0%
Chips	77.0	60.0	78.0%
Cigarettes cigars	77.0	57.0	74.0%
Chocolate tablets	72.0	29.0	40.0%
Toys	68.0	1.0	1.0%
Yogurt pudding	65.0	45.0	69.0%
Soft drinks	62.0	49.0	79.0%
Sauces	58.0	20.0	34.0%
Stationery	55.0	1.0	2.0%
Bacon cold cuts	51.0	36.0	71.0%

** The screenshot shows median values for each category.*

Commercial Team /
Sales Boost with Correlation Strategies

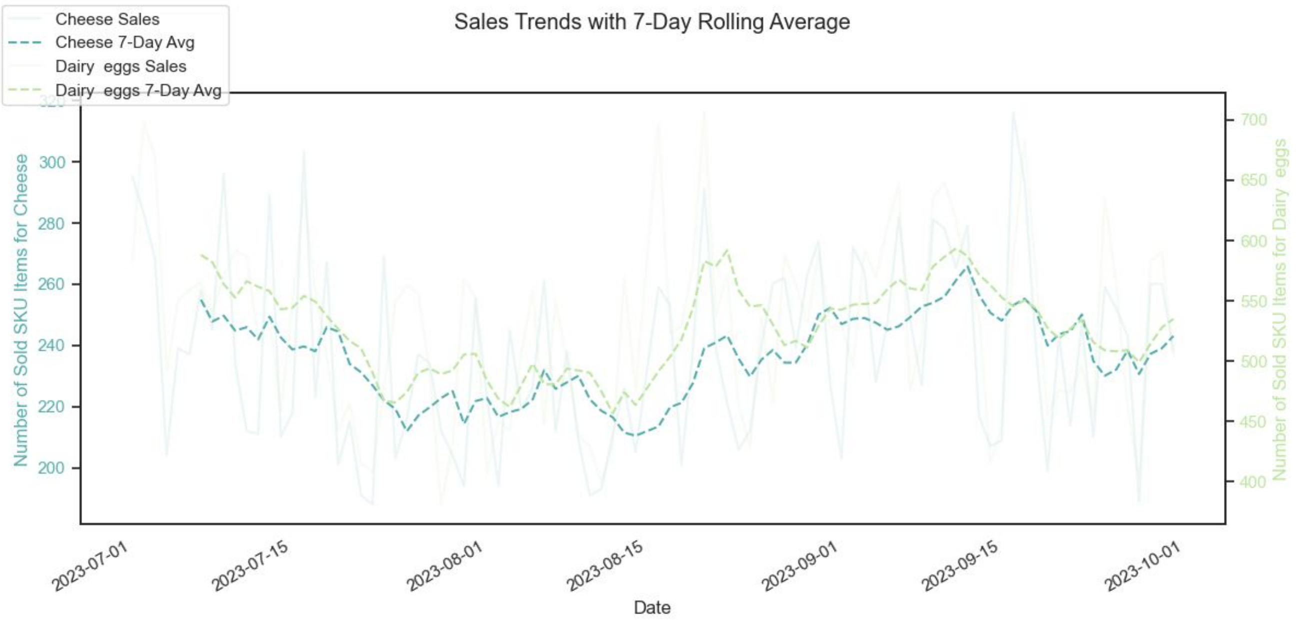
Commercial Team / Sales Boost with Correlation Strategies

Identifying correlations in sales trends across categories suggests revenue can be increased by strategically adjusting product recommendations and offerings.

Explore the interface for ways to refine recommendations or create more attractive options for customers, such as bundled promotions, leveraging demand correlations between categories.

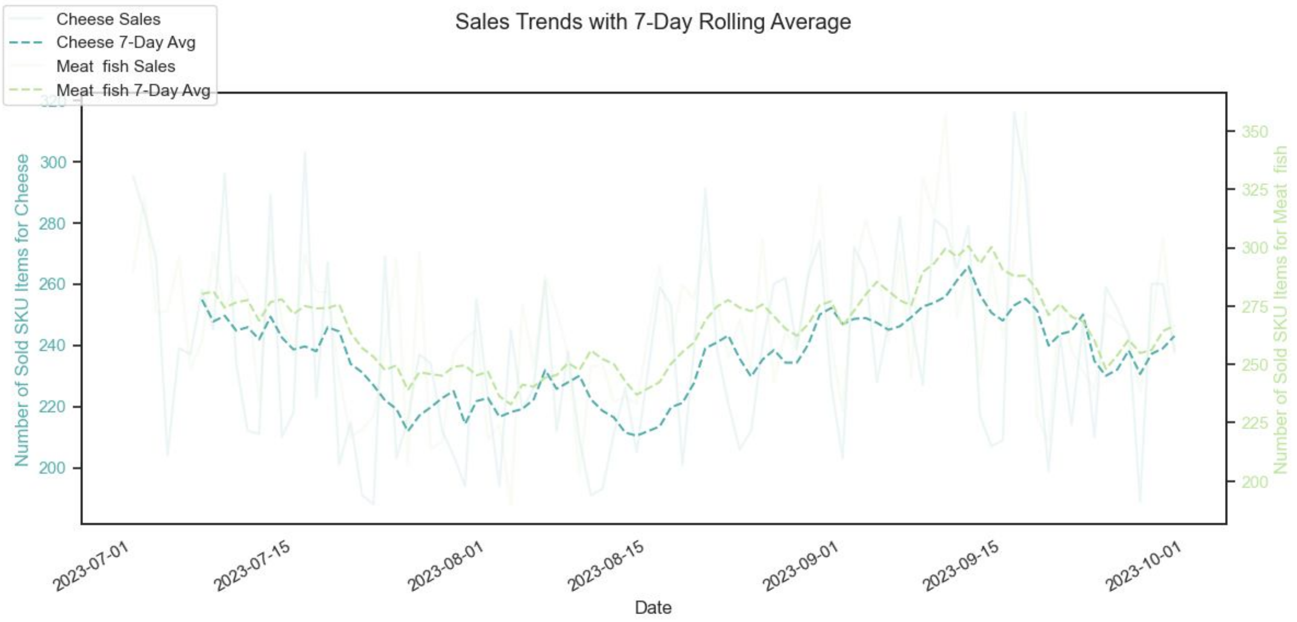
The next slides will present examples of these correlated categories.

Commercial Team / Sales Boost with Correlation Strategies



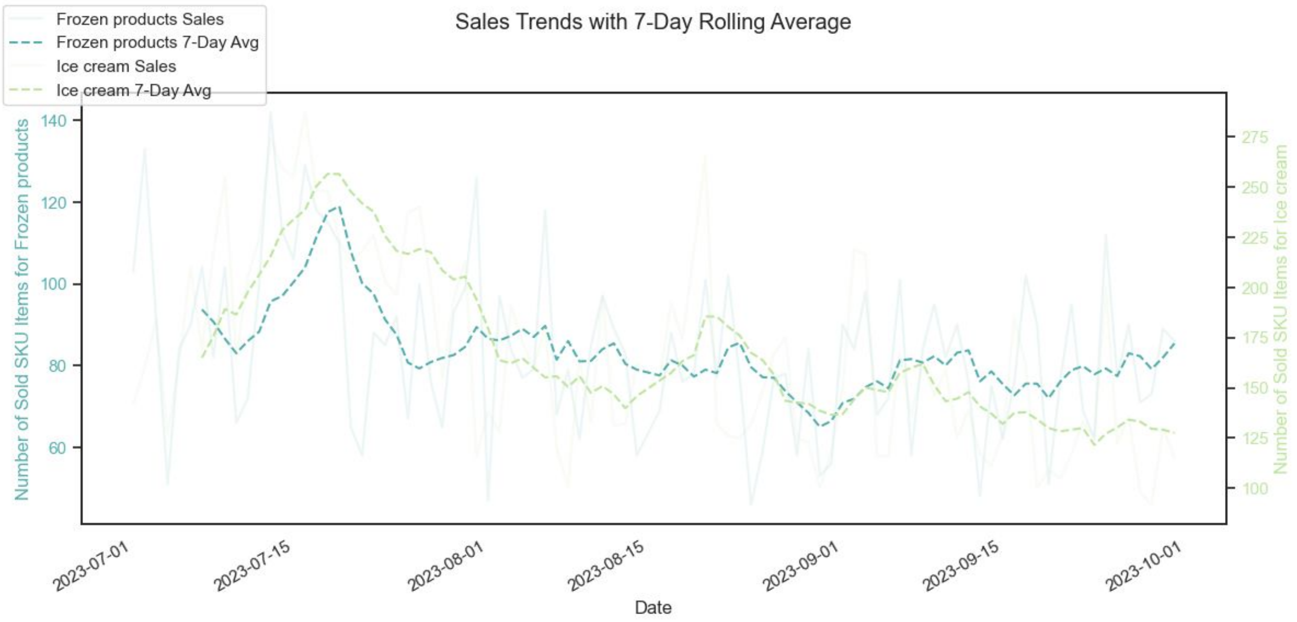
Cheese and Dairy eggs

Commercial Team / Sales Boost with Correlation Strategies



Cheese and Meat & Fish

Commercial Team / Sales Boost with Correlation Strategies



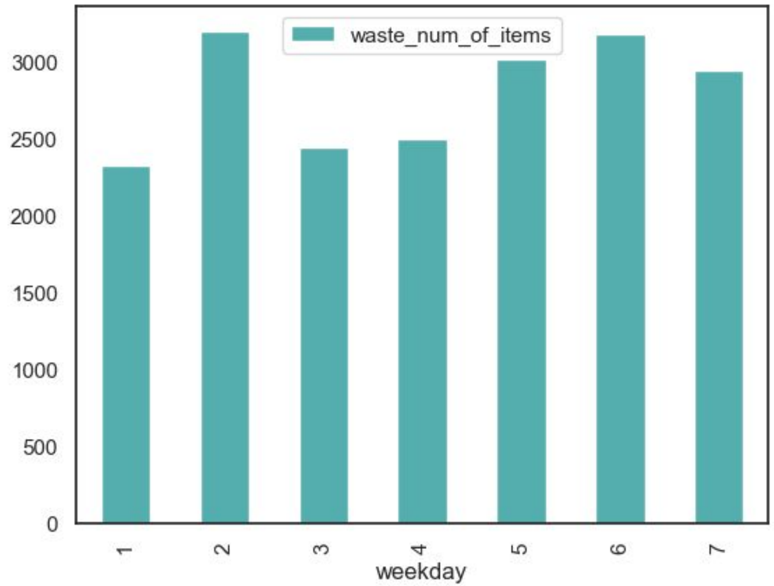
Frozen products and Ice cream

Supply chain /
Product Wastage: Analysis and Anomalies

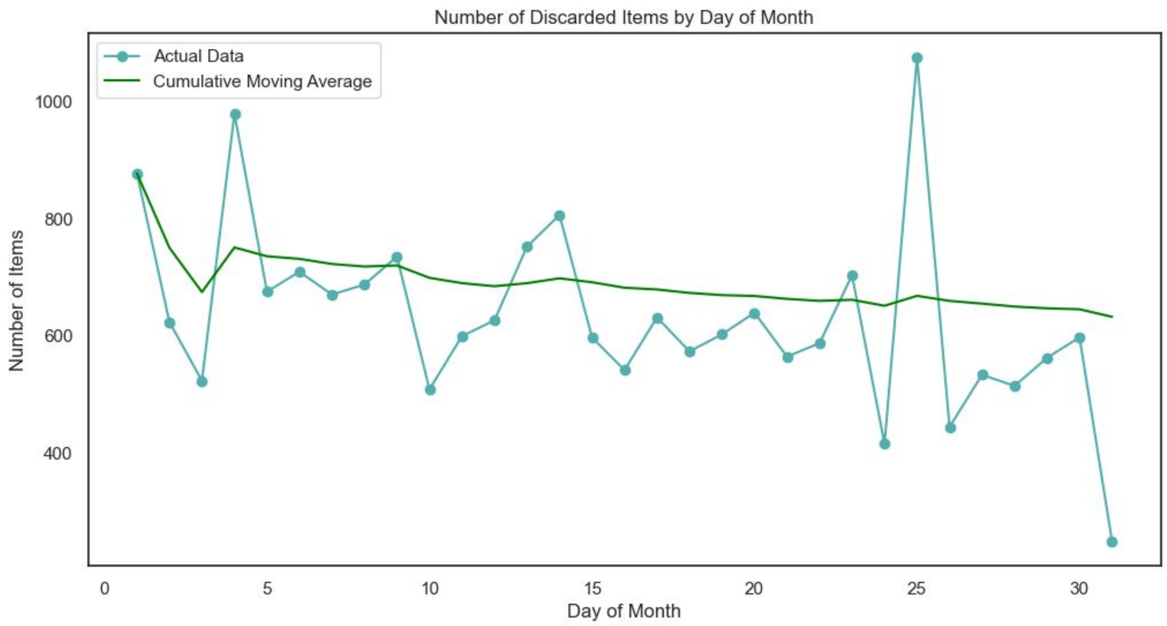
Supply chain / Product Wastage: Analysis and Anomalies

Examining data on wasted products reveals no consistent pattern across weekdays; the wastage is fairly uniform.

Here is the data for the top 10 categories with the highest wastage.

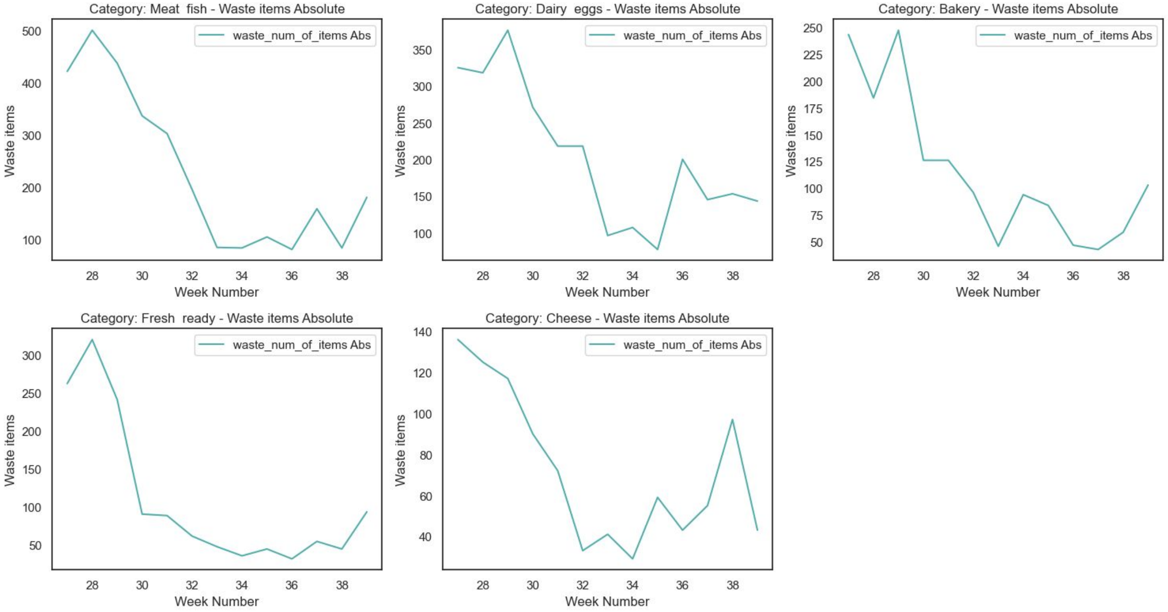


Supply chain / Product Wastage: Analysis and Anomalies



There's also no significant impact on wastage based on whether it's the end or beginning of the month.

Supply chain / Product Wastage: Analysis and Anomalies



However, there was a notable increase in wastage at the start of the period in several key categories, an anomaly that didn't recur in subsequent months. It's crucial to investigate these instances, collaborate with stores, and work towards reducing such occurrences.

Task 2, Orders info exploration

Using the provided dataset, please answer the questions shown below.

Questions are:

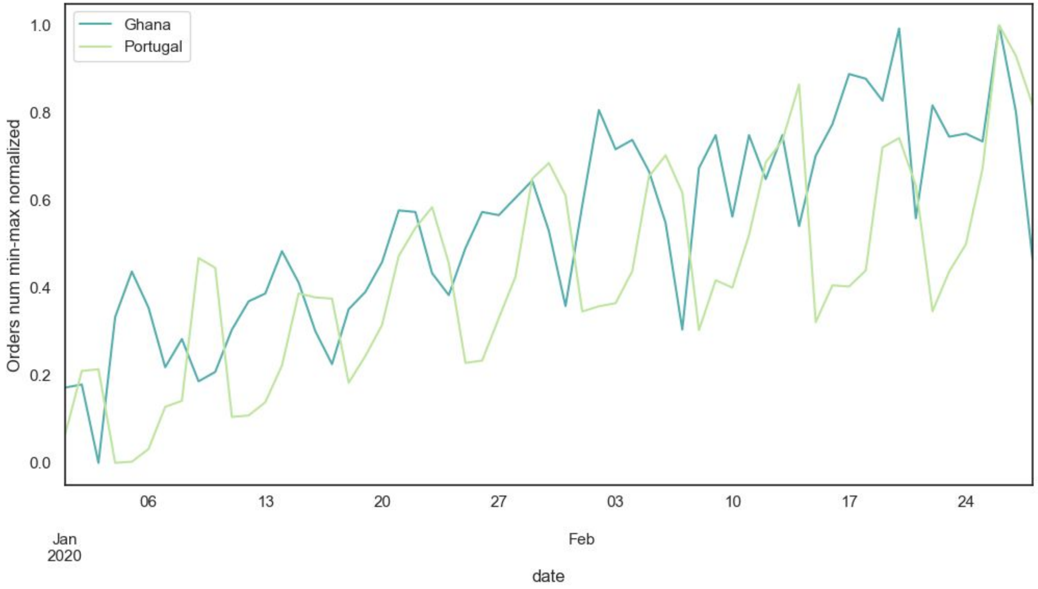
1) Do we have any seasonality in the countries shown?

2) Use your knowledge to predict with the available data, how many orders we will have in March 2020 in each country shown?

3) Please tell us any other valuable insight that you can extract from the data available and what would you do to solve it

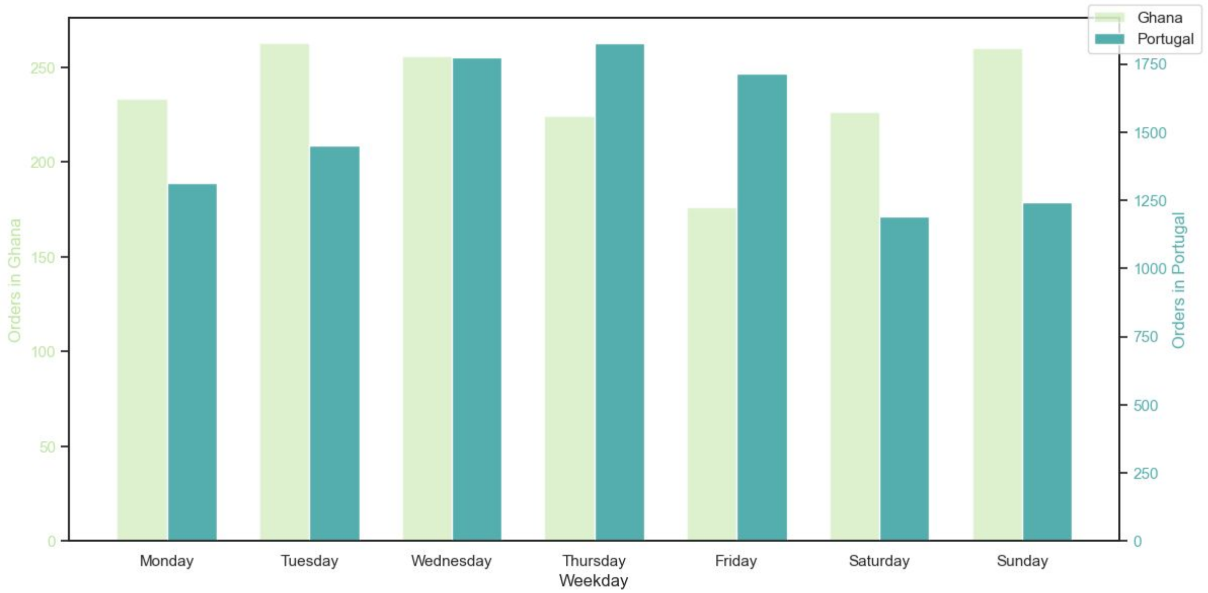
Do we have any seasonality in the countries?

Do we have any seasonality in the countries?



The data's timeframe is too short to confirm monthly seasonality but indicates a clear dependency on weekdays in both countries.

Do we have any seasonality in the countries?



In Ghana, order volumes begin increasing on Saturday, peaking on Tuesday. In Portugal, sales are higher from Wednesday to Friday.

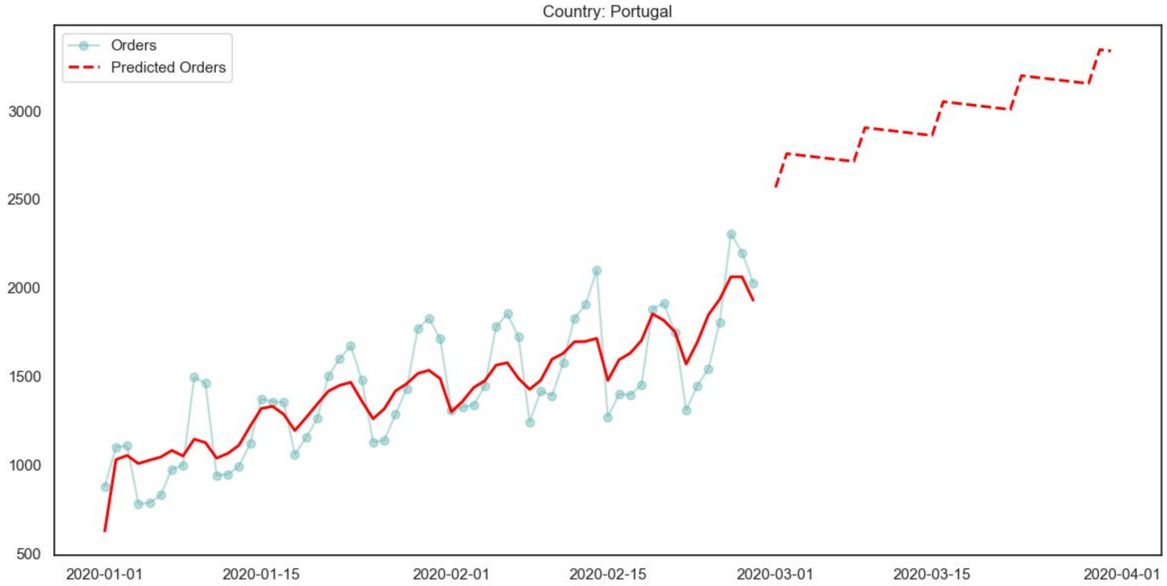
Predict how many orders
we will have in March 2020 in each country shown?

Predict how many orders we will have in March 2020 in each country shown?

The limited data volume and detail make it challenging to accurately model order predictions quickly.

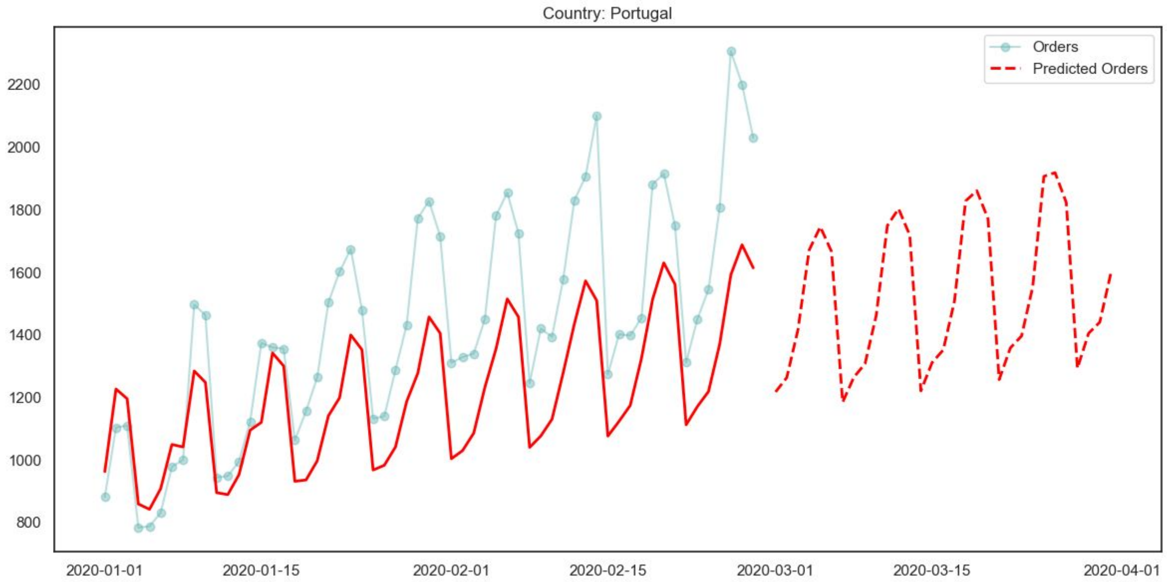
I'll present examples and a simplified short-term solution at this section’s conclusion.

Predict how many orders we will have in March 2020 in each country shown?



For example, with more features and data granularity, simple regression models overly optimistic predictions.

Predict how many orders we will have in March 2020 in each country shown?

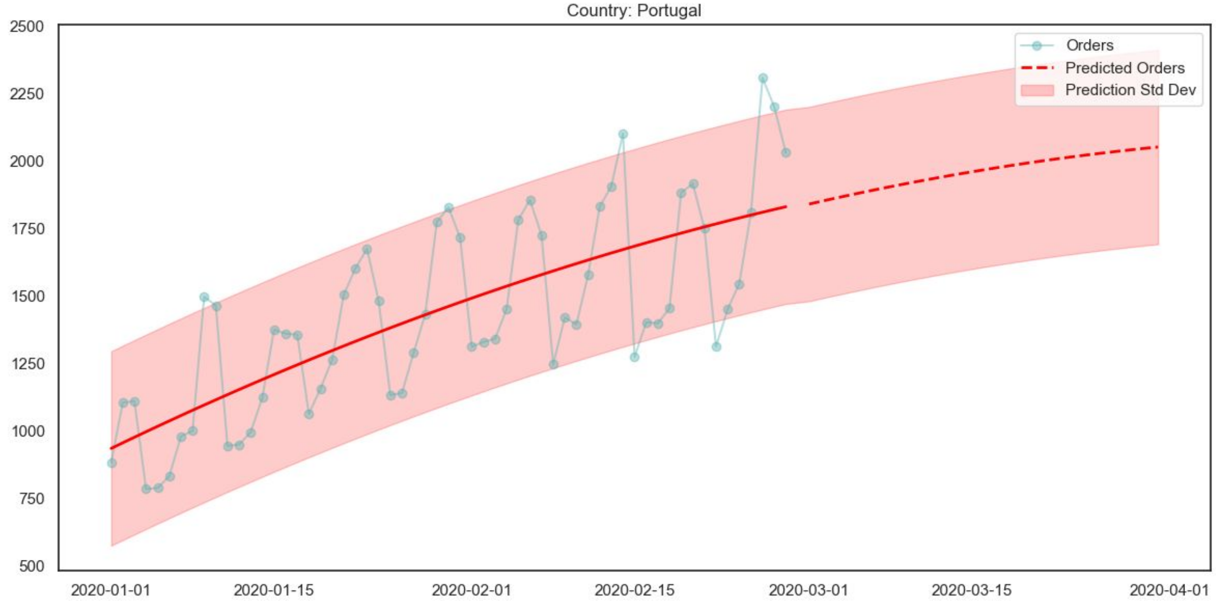


With fewer features (excluding specific restaurant data), the model becomes too pessimistic and diverges from actual data.

Predict how many orders we will have in March 2020 in each country shown?

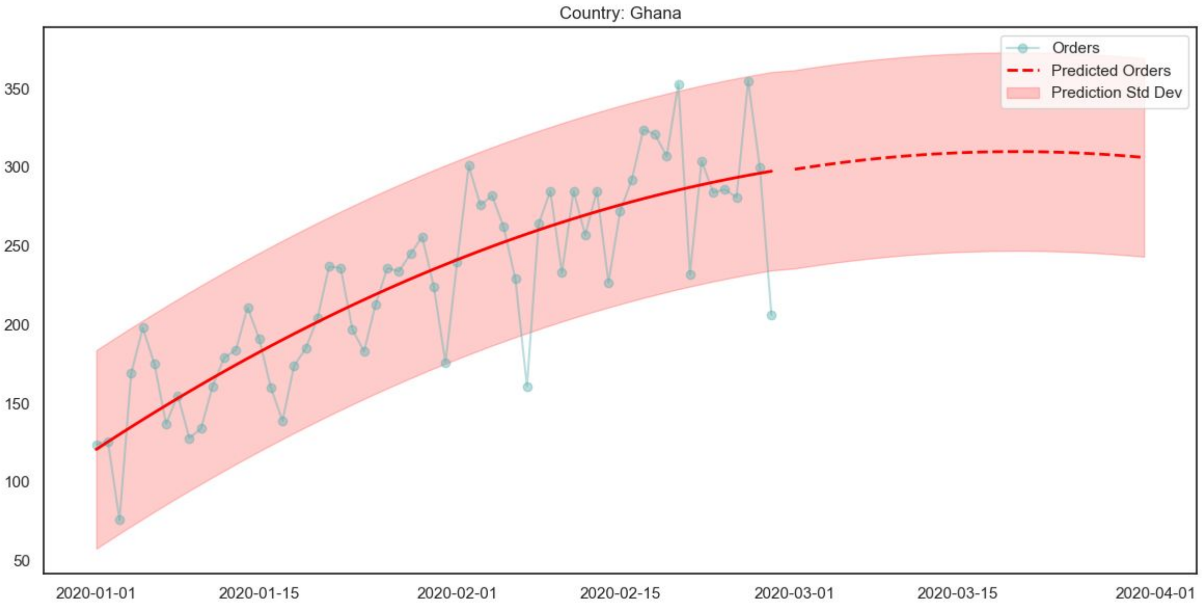
Given the data limitations, a feasible quick fix is using polynomial regression with standard deviation from the sample for short-term projections based on past trends.

Predict how many orders we will have in March 2020 in each country shown?



In Portugal, the model predicts March 2020 orders to range between 49 570 and 71 893.

Predict how many orders we will have in March 2020 in each country shown?



"In Ghana, March 2020 order projections range from 7 564 to 11 471.

Tell us any other valuable insight that you can extract from the data and what would you do to solve it

Tell us any other valuable insight that you can extract from the data and what would you do to solve it

Order cancellation practices vary significantly between Ghana and Portugal, with Ghana seeing more instances of restaurants requesting cancellations through customer support.

This could impact restaurant ratings and customer satisfaction.

Portugal

cancel_reason	cancellations	total	share
Could not find a courier to deliver the order	206	901	23%
The restaurant asked customer support to fail the order	50	901	6%
The restaurant rejected the order	628	901	70%
User cancellation	17	901	2%

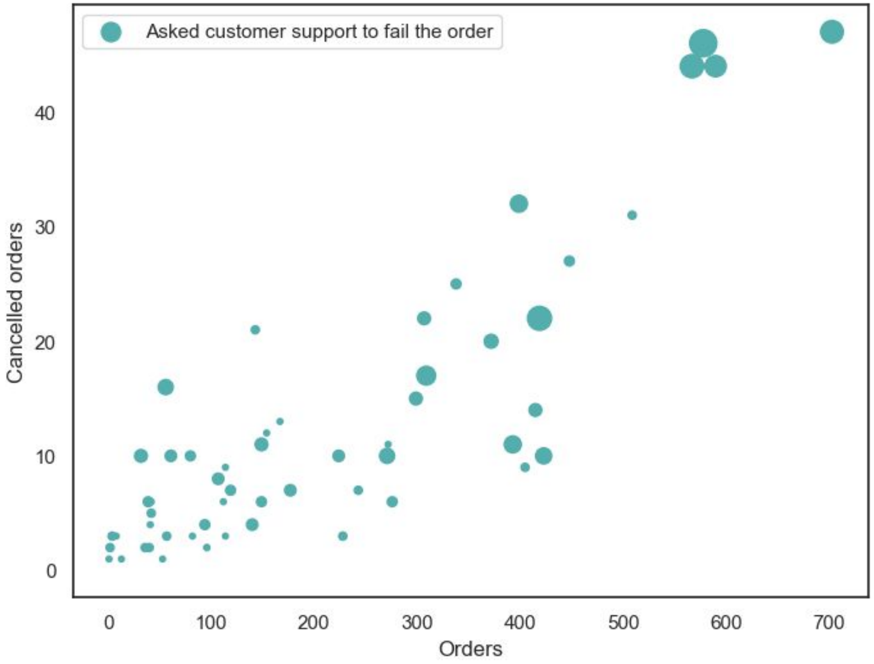
Ghana

cancel_reason	cancellations	total	share
Could not find a courier to deliver the order	146	781	19%
The restaurant asked customer support to fail the order	251	781	32%
The restaurant rejected the order	349	781	45%
User cancellation	35	781	4%

Tell us any other valuable insight that you can extract from the data and what would you do to solve it

As the restaurant receives more orders, the number of these cases rises. This trend has a few exceptions but is generally common in this market.

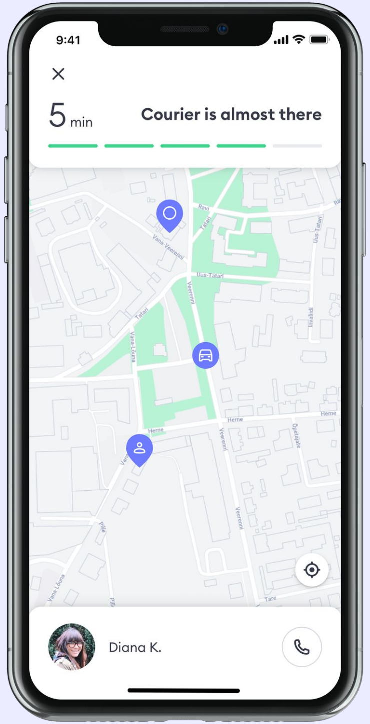
The next step is to seek further information from the market to understand why this is happening and to explore ways to adjust our scoring and recommendations systems accordingly.



Thank you!

It was interesting. What you saw is just a small part of the work done.

Please, review all the code and outputs in the attached notebooks :)



Ivan Kartavyi / Lisbon