

Homework task: Shipping prices

Context & Overview

Vinted has integrated shipping options for users to be able to seamlessly ship their items. We charge buyers for the shipping service based on the parcel size selected by the seller. Afterwards, we receive invoices from the shipping service providers.

Your task is to load the data in the provided format, transform it if necessary, and analyse it.

The task should take no longer than 4-8 hours. You can use any tools you find most convenient for the job.

Your final deliverables can be in your preferred format, but in addition to your conclusions we would like to see your work in a readable format, e.g. Jupyter Notebooks. You can place your findings/solution in any code sharing platform (Github, Bitbucket, etc.) or send it via email.

Assignment

1. In your analysis, please provide answers to the following questions:
 - Are there any discrepancies between the shipping service provider invoices and our invoices to buyers? Can you explain them?
 - What are the key trends you observe in the data over the period for which data is available? What is the business impact of these trends?
 - If you had more time and could ask for more data, how would you continue your analysis to identify and fix shortcomings in the shipping/billing process?
2. Please think about the production version of your solution and provide a structure of how the ETL pipeline could look like to arrive from the raw input data to the analytical reporting. Think of potential risks regarding the scalability and flexibility of your solution having the rapid growth of Vinted in mind.

Data

The provided data sets can be split into two:

- Has prefix “`provider_`” – data that we receive from third party service providers.
- Has prefix “`product_`” – data that we collect in Vinted.

A further explanation on what to expect from data in specific folders:

- `provider_invoices/*` - the shipping labels that the shipping provider is charging us for.
- `provider_prices/*` - the prices that we are being charged by the provider based on weight, route of the shipment.
- `product_shipments/*` - the list of shipments that we see in our data.
- `product_package_types/*` - the list of package types we have on our platform that a seller can select.
- `product_invoices/*` - the list of transactions where we have collected the shipment fee from the buyers.