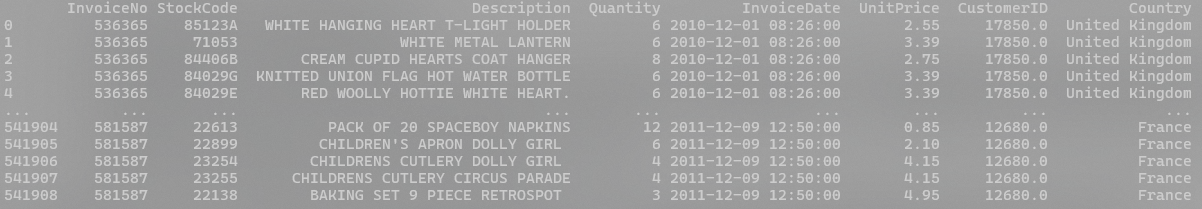
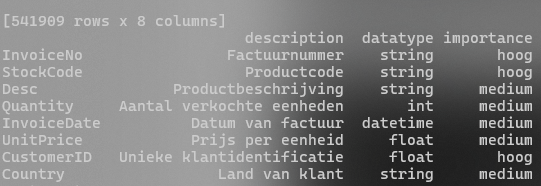
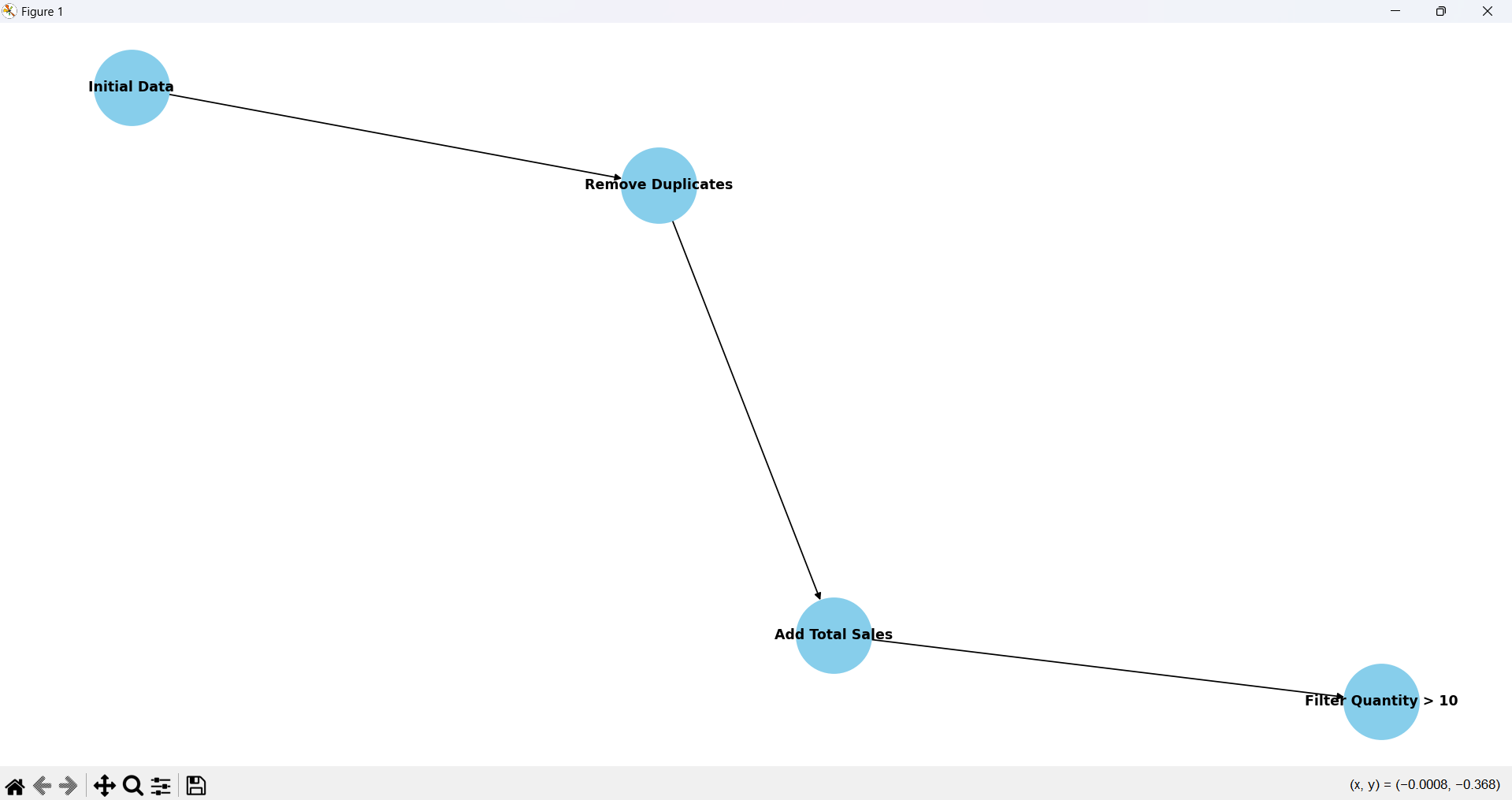
Learning**DataFoundations Labo 4**

**Kobe Vervoort 3ITSOF1**

Metadata & Data Lineage

Gegenereerde Data Catalogus: 

Overzicht Metadata: 

Data Lineage Grafiek:

De transformaties die de data heeft ondergaan, worden weergegeven in de data lineage grafiek. Zoals we kunnen zien op de grafiek werden volgende transformaties uitgevoerd op de data:

1. **Verwijder duplicaten**: Deze transformatie verwijderde rijen die identieke waarden bevatten voor alle kolommen. Het doel was om de dataset schoon te maken en de overbodige records eruit te halen.
2. **Toevoegen van de “total\_sales” kolom**: Deze kolom werd toegevoegd door de prijs per eenheid (UnitPrice) te vermenigvuldigen met het aantal verkochte eenheden (Quantity). Dit helpt bij het berekenen van de totale verkoopwaarde voor elke transactie.
3. **Filteren op Quantity > 10**: Deze transformatie filterde alle records uit waarin de hoeveelheid (Quantity) kleiner was dan 10. Dit werd gedaan om te focussen op de grotere sales, die relevanter zijn voor sales analyses.

Deze transformaties hadden telkens een andere **impact op de dataset**. Door duplicaten te verwijderen werd de dataset schoner, wat de betrouwbaarheid verhoogde. De toevoeging van de “total\_sales” kolom maakte het mogelijk om de omzet per transactie te berekenen. Het filteren van de dataset op grotere hoeveelheden zorgt voor een beter overzicht van de belangrijkste sales.

**Data lineage is belangrijk**, aangezien het helpt bij het traceren van elke stap in de transformatie van gegevens, waardoor de transparantie, controleerbaarheid en kwaliteit van de data verzekerd kan worden. Dit is cruciaal voor het maken van betrouwbare analyses en handig bij het trekken van conclusies (uit de data) en het maken van besluiten.