
Portfolio opdracht V

Transforming Data

BIMLEU04R2 – Business Intelligence
Onderwijs Periode 2 studiejaar 2025-2026

Portfolio opdracht week 5 – Transforming Data

Opdracht

Casus: Eventura Analytics

Eventura Analytics is een middelgroot Nederlands data-adviesbureau dat zich specialiseert in datagedreven besluitvorming voor de evenementenbranche. Het bureau werkt voor gemeenten, organisatoren van festivals en commerciële evenementenbureaus die inzicht willen krijgen in bezoekersstromen, risico's en prestaties van evenementen in de openbare ruimte.

Situatie

De afgelopen jaren is het organiseren van evenementen complexer geworden. Weersomstandigheden spelen een steeds grotere rol in de opkomst van bezoekers, de inzet van beveiliging en de uiteindelijke winstgevendheid van evenementen. Gemeenten en organisatoren vragen Eventura Analytics daarom om beter te onderbouwen welke invloed het weer daadwerkelijk heeft op bezoekersaantallen en hoe zij hier in de toekomst rekening mee kunnen houden bij vergunningverlening en planning.

Eventura beschikt over meerdere databronnen, afkomstig uit verschillende systemen en externe partijen. Deze data is echter niet direct geschikt voor analyse. Bronnen verschillen in structuur, datumnotaties en definities. Daarnaast blijkt dat externe weerdata van verschillende leveranciers tegenstrijdige informatie kan bevatten over exact dezelfde dag en locatie. In de praktijk betekent dit dat er geen eenduidige "waarheid" is, maar dat een BI-analist keuzes moet maken op basis van data-analyse en onderbouwing.

Jouw rol en opdracht

Als junior BI-analist bij Eventura Analytics krijg jij de opdracht om deze ruwe data om te zetten naar een betrouwbare, analyseerbare data die gebruikt kan worden voor dashboards en beleidsanalyses. Daarbij wordt van je verwacht dat je niet alleen technisch data opschoont en samenvoegt, maar ook kritisch nadenkt over datakwaliteit. Hoe kan ik betrouwbaarheid van de set duiden? Waar zie je afwijkingen? En welke aannames maak je daarbij?

In deze opdracht staat daarom niet alleen de techniek centraal, maar vooral jouw rol als BI-professional die data begrijpt, beoordeelt en vertaalt naar bruikbare informatie voor besluitvorming.

Data bestanden en EDA

In deze stap onderzoek je welke data je hebt, waar de verschillen zitten en welke bron het meest betrouwbaar lijkt. Eventura Analytics heeft je één grote data set gegeven en deze set is samengesteld uit verschillende operationele databases van Eventura Analytics.

Events.csv: dit bestand bevat informatie over georganiseerde evenementen door Eventura Analytics aangevuld met verschillende informatievelden die het weer van die dag beschrijven.

Opdracht 1.1) Maak een nieuw python bestand (.py) aan en lees het CSV-bestand in met Panda's in Python zodat je er een dataframe van maakt. Hoe je dit doet is behandeld in lesweek 2.5.

Opdracht 1.2) Voer nu in python een Exploratory Data Analysis uit op de dataset. Je gaat dit doen met behulp van Python! Je voert in deze stap nog geen definitieve transformaties uit, maar analyseert de data kritisch zoals een BI-analist dat zou doen. Vul jouw eerder gemaakte Python code (van opdracht 1.1) aan en schrijf code waarmee je aan het einde van je EDA je tenminste antwoord hebt op onderstaande vragen:

1. Uit hoeveel rijen en kolommen bestaat de dataset?
2. Hoeveel procent van de datavelden in elke kolom is leeg?
3. Hoeveel dubbele regels zitten er in de dataset?
4. Hoeveel lege regels zitten er in de dataset?
5. Welke verschillende data formaten worden er gebruikt voor datum kolommen?
6. Welke overeenkomsten en verschillen tussen kolomnamen zie je? Wat betekenen de kolommen?
7. Zie je kolommen waarbij de data afwijkt bij wat je verwacht?

Verzin naast deze vragen zelf nog minimaal twee vragen die je beantwoord tijdens je Exploratory Data Analysis.

Opdracht 1.3) Presenteer Eventura Analytics jouw bevindingen over de kwaliteit van de dataset. Reflecteer over de impact van jouw bevindingen. Welke kolommen zijn op basis van jouw analyse het meest betrouwbaar en welke juist niet? Welke aannames maak je hierbij? Er is geen standaard goed of fout antwoord, onderbouwd je keuzes.

Ontwerp een eigen methode om de datakwaliteit te beoordelen conform BIV-classificatie. Welke score van 0-10 zou je de dataset nu geven?

Werk jouw antwoorden en bevindingen uit opdracht 1.3 visueel uit in een AS-IS creatieve visuele samenvatting. Denk hierbij aan een infographic, schematische tekening. Dit mag in een bestandsformaat naar keuze (bijvoorbeeld .pdf, .docx, .jpeg, .pbix, etc.) De visual moet laten zien welke data je hebt, waar de verschillen zitten, welke keuzes jij maakt en welke score je de dataset geeft. Zorg dat jouw studentennummer en naam op de uitwerking te zien zijn.

EDA in Actie!

Nu we de data verkend en onderbouwd hebben gaan we deze inzichten omzetten naar actie. We gaan de dataset opschonen en transformeren zodat ze klaar zijn voor het importeren in het datawarehouse.

Opdracht 2.1) Werk verder in het eerder aangemaakte python bestand van opdracht 1. Schrijf python code waarmee je de datakwaliteit van de dataset gaat verhogen. Welke stappen je precies uitvoert is afhankelijk van jouw bevindingen en besluiten die je in je Exploratory Data Analysis hebt gemaakt. Zorg dat je een nieuw .csv bestand wegschrijft met de opgeschoonde dataset.

Opdracht 2.2) Kijk nu nogmaals naar de scoringsmethode die je zelf hebt verzonnen. Welke score krijgt de dataset na opschoning & transformaties? Werk de antwoorden en bevindingen uit opdrachten 2.1 en 2.2 visueel uit in TO-BE creatieve visuele samenvatting. Dit mag in een bestandsformaat naar keuze (bijvoorbeeld .pdf, .docx, .jpeg, .pbix, etc.), maar moet aansluiten op je AS-IS samenvatting zodat je deze gemakkelijk kunt vergelijken. De visual moet laten zien welke acties je hebt uitgevoerd in jouw transformaties en de resultaten hiervan. Neem op deze uitwerking een korte reflectie **Zorg dat jouw studentennummer en naam op de uitwerking te zien zijn.**

Je hebt nu een opgeschoonde, nette dataset! Hier kan Eventura Analytics weer mee verder. Goed gedaan!

Wat lever je in voor je Portfolio?

Twee bestanden:

- Één werkbaar python bestand met als bestandsnaam:

`Portfolio_opdr5_[jouw studentnummer].py`

`Voorbeeld: Portfolio_opdr5_12737848.py`

Let op: Hierin zit de code voor het inlezen, analyseren én opschonen van de dataset.

Deze code moet werken met het bestand wat aangeleverd wordt.

- Een visuele uitwerking van jouw AS-IS situatie in bestandvorm naar keuze.

- Een visuele uitwerking van jouw TO-BE situatie mét **reflectie** in bestandvorm naar keuze. In je reflectie Neem in de reflectie de volgende vragen mee:

Hoe heb je het proces van analyse en data cleanisng aangepakt?
Wat ging goed in jouw aanpak, en wat zou je de volgende keer anders doen?