**Dataset Information**

1. **Disclaimer**

The Look adalah situs pakaian eCommerce fiktif yang dikembangkan oleh tim [Looker](https://cloud.google.com/looker?). Dataset berisi informasi tentang pelanggan, produk, pesanan, logistik, acara web, dan kampanye pemasaran digital. Isi dari kumpulan data ini adalah sintetis, dan diberikan kepada praktisi industri untuk tujuan penemuan, pengujian, dan evaluasi produk.

1. **Source**

Terdapat dua rujukan sumber data yang dapat dilihat di link berikut ini:

Google Big Query: The Look Ecommerce

<https://github.com/GoogleCloudPlatform/public-datasets-pipelines/blob/main/datasets/thelook_ecommerce/pipelines/_images/run_thelook_kub/fake.py>

Kaggle: Absenteeism Dataset

<https://www.kaggle.com/datasets/HRAnalyticRepository/absenteeism-dataset>

1. **Dataset**

Dataset terbagi menjadi dua kategori yaitu dataset utama (*primary*) dan dataset pendukung (*secondary*). Dataset ini akan dikelompokkan berdasarkan penggunaan dan kebutuhan analitik di tiap departemen.

1. **Goals**

Anda diminta untuk membuat :

| List Task | Output | Mandays |
| --- | --- | --- |
| Design Data Warehouse | Design, Script | 2 |
| Design Data Lake | Design | 1 |
| Service Ingestion to Data Lake | Script | 1 |
| Service Tranformation | Script | 1 |
| Service Load to Data Warehouse | Script | 1 |

1. **Dataset Utama (Primary Dataset)**
   1. Data Users (Departemen Marketing)

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| id | INTEGER | NULLABLE |
| first\_name | STRING | NULLABLE |
| last\_name | STRING | NULLABLE |
| email | STRING | NULLABLE |
| age | INTEGER | NULLABLE |
| gender | STRING | NULLABLE |
| state | STRING | NULLABLE |
| street\_address | STRING | NULLABLE |
| postal\_code | STRING | NULLABLE |
| city | STRING | NULLABLE |
| country | STRING | NULLABLE |
| latitude | FLOAT | NULLABLE |
| longitude | FLOAT | NULLABLE |
| traffic\_source | STRING | NULLABLE |
| created\_at | TIMESTAMP | NULLABLE |

| Field name | Description |
| --- | --- |
| id | ID User |
| first\_name | Nama depan User |
| last\_name | Nama belakang User |
| email | Email User |
| age | Usia User |
| gender | Jenis kelamin User |
| state | Asal negara bagian User |
| street\_address | Alamat User |
| postal\_code | Kode pos user |
| city | Asal kota User |
| country | Asal negara User |
| latitude | Titik latitude |
| longitude | Titik longitude |
| traffic\_source | Asal User mendapati situs ini |
| created\_at | Waktu yang tercatat saat User membuat akun |

* 1. Data Order\_items (Departemen Sales)

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| id | INTEGER | NULLABLE |
| order\_id | INTEGER | NULLABLE |
| user\_id | INTEGER | NULLABLE |
| product\_id | INTEGER | NULLABLE |
| inventory\_item\_id | INTEGER | NULLABLE |
| status | STRING | NULLABLE |
| created\_at | TIMESTAMP | NULLABLE |
| shipped\_at | TIMESTAMP | NULLABLE |
| delivered\_at | TIMESTAMP | NULLABLE |
| returned\_at | TIMESTAMP | NULLABLE |
| sale\_price | FLOAT | NULLABLE |

| Field name | Description |
| --- | --- |
| id | ID order items |
| order\_id | ID Order (tabel) |
| user\_id | ID User (tabel) |
| product\_id | ID Product (tabel) |
| inventory\_item\_id | ID Inventory Item (tabel) |
| status | Status Order |
| created\_at | Waktu yang dicatat saat order dipesan |
| shipped\_at | Waktu yang dicatat saat order dikemas |
| delivered\_at | Waktu yang dicatat saat order dikirim |
| returned\_at | Waktu yang dicatat saat order dikembalikan oleh User |
| sale\_price | Harga jual barang ($) |

1. Data Inventory Items atau Events (Departemen Product)
   1. Berikut adalah tabel Inventory Items berisi daftar nama-nama kolom beserta dengan informasi tipe data.

|  |  |  |
| --- | --- | --- |
| Field name | Data Type | Constraint |
| id | INTEGER | NULLABLE |
| product\_id | INTEGER | NULLABLE |
| created\_at | TIMESTAMP | NULLABLE |
| sold\_at | TIMESTAMP | NULLABLE |
| cost | FLOAT | NULLABLE |
| product\_category | STRING | NULLABLE |
| product\_name | STRING | NULLABLE |
| product\_brand | STRING | NULLABLE |
| product\_retail\_price | FLOAT | NULLABLE |
| product\_department | STRING | NULLABLE |
| product\_sku | STRING | NULLABLE |
| product\_distribution\_center\_id | INTEGER | NULLABLE |

|  |  |
| --- | --- |
| Field name | Description |
| id | ID Inventory Items |
| product\_id | ID product (tabel) |
| created\_at | Waktu produk dibuat |
| sold\_at | Waktu produk terjual |
| cost | Harga asli produk |
| product\_category | Kategori produk |
| product\_name | Nama produk |
| product\_brand | Merek produk |
| product\_retail\_price | Harga retail produk |
| product\_department | Departemen produk (untuk laki-laki/perempuan) |
| product\_sku | SKU Produk |
| product\_distribution\_center\_id | ID Distribution center (tabel) |

* 1. Berikut adalah tabel Events berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| id | INTEGER | NULLABLE |
| user\_id | INTEGER | NULLABLE |
| sequence\_number | INTEGER | NULLABLE |
| session\_id | STRING | NULLABLE |
| created\_at | TIMESTAMP | NULLABLE |
| ip\_address | STRING | NULLABLE |
| city | STRING | NULLABLE |
| state | STRING | NULLABLE |
| postal\_code | STRING | NULLABLE |
| browser | STRING | NULLABLE |
| traffic\_source | STRING | NULLABLE |
| uri | STRING | NULLABLE |
| event\_type | STRING | NULLABLE |

| Field name | Description |
| --- | --- |
| id | ID events |
| user\_id | ID user (tabel) |
| sequence\_number | Urutan navigasi yang dilakukan User dalam 1 sesi |
| session\_id | ID Session dari *event* ketika User membuka website |
| created\_at | Waktu yang tercatat saat User memulai *event* |
| ip\_address | IP address akses |
| city | Kota akses |
| state | Negara bagian akses |
| postal\_code | Postal code akses |
| browser | Browser yang digunakan |
| traffic\_source | Asal User mendapati situs ini |
| uri | Informasi mengenai Url mengenai *page* yang diakses Users (selain *homepage*) |
| event\_type | Kategori *page* yang diakses User |

1. Data Employees (Departemen Human Resources)

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

|  |  |  |
| --- | --- | --- |
| Field name | Data Type | Constraint |
| Fisrt\_Name | STRING | NULLABLE |
| Last\_Name | STRING | NULLABLE |
| Gender | FLOAT | NULLABLE |
| Age | FLOAT | NULLABLE |
| Length\_Service | FLOAT | NULLABLE |
| Absent\_Hours | FLOAT | NULLABLE |
| distribution\_centers\_id | INTEGER | NULLABLE |

|  |  |
| --- | --- |
| Field name | Description |
| Fisrt\_Name | Nama depan pegawai |
| Last\_Name | Nama belakang pegawai |
| Gender | Jenis kelamin pegawai |
| Age | Usia pegawai |
| Length\_Service | Lama pegawai bekerja diperusahaan The looker (tahun) |
| Absent\_Hours | Lama absen dalam satu tahun (jam) |
| distribution\_centers\_id | Distribution centers id (tabel) |

1. **Dataset Pendukung (Secondary Dataset)**
   1. Data Distribution\_Centers

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| id | INTEGER | NULLABLE |
| name | STRING | NULLABLE |
| latitude | FLOAT | NULLABLE |
| longitude | FLOAT | NULLABLE |

| Field name | Description |
| --- | --- |
| id | Id distribution centers |
| name | Nama Distribution Centers |
| latitude | Titik koordinat latitude |
| longitude | Titik koordinat longitude |

* 1. Data Orders

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| order\_id | INTEGER | NULLABLE |
| user\_id | INTEGER | NULLABLE |
| status | STRING | NULLABLE |
| gender | STRING | NULLABLE |
| created\_at | TIMESTAMP | NULLABLE |
| returned\_at | TIMESTAMP | NULLABLE |
| shipped\_at | TIMESTAMP | NULLABLE |
| delivered\_at | TIMESTAMP | NULLABLE |
| num\_of\_item | INTEGER | NULLABLE |

| Field name | Description |
| --- | --- |
| order\_id | ID order |
| user\_id | ID User (tabel) |
| status | Status order |
| gender | Jenis kelamin User |
| created\_at | Waktu Order dibuat oleh user |
| returned\_at | Waktu order yang dikembalikan oleh user |
| shipped\_at | Waktu order yang dikemas oleh user |
| delivered\_at | Waktu order yang dikirim oleh user |
| num\_of\_item | Jumlah item yang dipesan |

* 1. Data Products

Berikut adalah tabel berisi daftar nama-nama kolom beserta dengan informasi tipe data.

| Field name | Data Type | Constraint |
| --- | --- | --- |
| id | INTEGER | NULLABLE |
| cost | FLOAT | NULLABLE |
| category | STRING | NULLABLE |
| name | STRING | NULLABLE |
| brand | STRING | NULLABLE |
| retail\_price | FLOAT | NULLABLE |
| department | STRING | NULLABLE |
| sku | STRING | NULLABLE |
| distribution\_center\_id | INTEGER | NULLABLE |

| Field name | Description |
| --- | --- |
| id | ID Products |
| cost | Harga produk saat produksi |
| category | Kategori dari produk |
| name | Nama dari produk |
| brand | Brand dari produk |
| retail\_price | Harga retail |
| department | Departemen produk (untuk laki-laki/perempuan) |
| sku | SKU produk |
| distribution\_center\_id | Id distribution center (tabel) |