|  |  |  |
| --- | --- | --- |
| **RETRIEVE SQL DATA DAN EKSPLORASI** | | |
| Praktikan | **Aslab** | |
| Nama: xxxx  Nim: xxxx | Annur Hangga Prihadi | 065001800028 |
| Faiz Kumara | 065001800003 |

**PRAKTIKUM 1**

**DATA SAINS DAN ANALITIK**

Pertemuan praktikum pertama kali ini adalah memasang dan menyesuaikan perangkat lunak maupun library yang dibutuhkan saat praktikum, mengetahui cara untuk mengambil data dari RDBMS, dan eksplorasi data sederhana dari data yang telah diambil oleh praktikan.

**Latihan 1**

**Pre-praktikum**

|  |
| --- |
| **Pastikan sudah mengunduh file orders.csv** |
|  |
| **Buka pgAdmin (Versi berapapun bisa)** |
|  |

|  |
| --- |
| **Buat database dengan nama dsda dengan cara klik kanan pada icon database > create > database** |
|  |

|  |
| --- |
| **Buat tabel dengan cara memperluas database dsda > schemas > public > tables > klik kanan pada tables > create > table** |
|  |
| **Beri nama table bebas terserah kalian, lalu klik columns** |
| **Tambah kolom dan tipe data sesuai dengan data order** |
|  |
| **Import data csv yang telah diunduh di-step pertama dengan cara klik kanan tabel yang baru dibuat > import/export** |
|  |
| **Sebelum import pastikan dahulu delimiter/pembatas dari data csv yang diunduh dengan cara klik kanan file > open with > notepad** |
| **Terlihat di dalam file orders milik aslab terdapat delimiter berupa tanda ; (Masing-masing praktikan kemungkinan mendapatkan delimiter yang berbeda, jadi tolong lebih teliti)** |
| **Setelah mengetahui delimiter atau pembatas lanjut ke tahap import file, lakukan langkah berikut**   1. **Klik export untuk berubah ke import** 2. **Pilih directory file penyimpanan file csv yang kalian unduh di-step awal** 3. **Pilih yes pada header** 4. **Sesuaikan delimiter dengan yang kalian dapat di-step sebelumnya** 5. **Jika mengikuti step akan muncul notifikasi sukses** |
|  |

**Memasang library yang dibutuhkan**

|  |
| --- |
| import sys  !{sys.executable} -m pip install psycopg2  !{sys.executable} -m pip install pandas  import psycopg2  import pandas as pd |
|  |

**Menyambungkan koneksi worksheet jupyter dengan PostgreSQL**

|  |
| --- |
| conn = psycopg2.connect(host="localhost", port = xxxx, database="dsda", user="postgres", password="xxxx")  conn.set\_session(autocommit=True)  cur = conn.cursor()  sql = "SELECT \* FROM public.nama\_tabel"  cur.execute(sql)  baris = cur.fetchall()  **--Sesuaikan port, password, dan nama\_tabel dengan milik kalian--** |
|  |

**Menampilkan n data pertama**

|  |
| --- |
| order = pd.DataFrame([[ij for ij in i]for i in baris])  order.rename(columns={0:'order\_id',1:'customer\_id',2:'campaign\_id',3:'order\_date',4:'city',5:'state',6:'zipcode',7:'payment\_type',8:'total\_price',9:'num\_order',10:'num\_units'},inplace=True)  order.head(4) |
|  |

**Menampilkan rangkuman statistik dari data integer dan dimensi tabel**

|  |
| --- |
| order.describe()  order.shape |
|  |

**Menampilkan hasil dari rata-rata dan simpangan baku 1 kolom maupun berdasarkan kelompok**

|  |
| --- |
| order['total\_price'].mean()  order.groupby('order\_date')['total\_price'].mean() |
|  |

|  |
| --- |
| order['total\_price'].std()  order.groupby('order\_date')['total\_price'].std() |
|  |

**Latihan 2**

1. Gunakan data orders untuk mengetahui pengaruh data kolom lain terhadap total harga dari pesanan.
2. Berapakah nilai rata-rata dari total harga untuk tiap kategori campaign\_id, city, payment\_tipe dan Jelaskan makna dari keterkatian tiap kategori tersebut terhadap variabel total harga!
3. Berapakah nilai standar deviasi dari total harga untuk tiap kategori campaign\_id, city, payment\_tipe dan Jelaskan makna dari keterkatian tiap kategori tersebut terhadap variabel total harga!

**Lampiran Screenshot hasil a dan b**

|  |
| --- |
| **Input screenshot disini** |

**Makna dari hasil nilai rata-rata dan standar deviasi**

|  |
| --- |
| **Ketik makna disini** |