

STUDI INDEPENDEN BERSERTIFIKAT
E-COMMERCE

**Informasi Tugas dan Kertas Tugas Capstone Project
TM021-IMR-TI dan Data-Tugas02**

Bagian A. Informasi Tugas

1. Untuk Kelompok Tugas
 - a. Isi data di kotak berwarna hijau
 - b. Setelah tugas dikerjakan, unggah dokumen secara utuh (Bagian A, B, C, D jangan dihapus) ke:
 - i. Draft Kertas Tugas ke GDrive folder “03 Kertas Tugas-draft”
 - ii. Kertas Tugas Final ke GDrive folder “04 Kertas Tugas-final”
 - c. Cara penamaan Kertas Tugas: No Urut Tugas – Draft/Final – Kode Kelompok – Kode Sub Kelompok (bila ini tugas bersama isi dengan “Z”)-Nama Kelompok. Contoh:
 - i. Tugas01-Draft-BT02-01-A-Beauty01-Amethyst
 - ii. Tugas02-Final-BT02-05-Z-FnB01

2. Data Tugas

ID Tugas	TM021		Nama Mentor Pembimbing Tugas	Muhammad Imran
No Urut Tugas	Tugas02		Tipe Kelompok	Tipe B
Hari dan Tanggal Tugas	Sabtu. 20 November 2021		Bidang / Profesi	IT dan Data
Hari dan Tanggal Kertas Tugas diserahkan ke Kelompok Tugas	Sabtu. 20 November 2021		Catatan tambahan	Rubrik penilaian masih dalam proses penyusunan oleh mentor pembimbing dan mentor penilai
Nama Mentor Penilai Tugas (MNIT)	Valentinus Paramartha		Tanggal penyerahan draft Kertas Tugas dan PPT ke MNIT	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas)

Target tanggal penyerahan draft	Sabtu, 27 November 2021	Tanggal Mentoring Tugas Sinkronus	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas)
Tanggal penilaian Kertas Tugas dan PPT	(diisi oleh MNIT, diisi sesuai data di GSheet Sentra Tugas)	Tanggal penyerahan Kertas Tugas dan PPT final ke MNIT	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas)
Data Kelompok Tugas			
Kode Kelompok Tugas dan Nama Kelompok Tugas	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas)	Nama Sub-Kelompok Tugas	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas)
Nama Anggota Sub-Kelompok yang berkontribusi	(diisi oleh Kelompok Tugas, sesuai data di GSheet Sentra Tugas) 1. Arif Widagdo	Nama Anggota Kelompok yang berkontribusi dalam pengerjaan tugas / review tugas	(diisi oleh Kelompok Tugas) 1.
	2. Eko Prasetyo	Nama Anggota Kelompok yang berkontribusi dalam pengerjaan tugas / review tugas	2.
	3. Harnum Gina Fortuna		3.
	4. Putra Surya Jaya Togatorp		4.
			5.
			6.
			7.
			8.
			9.
			10.

3. Uraian Tugas

Automatisasi data pipelines dengan menggunakan python atau proses data ingestion menggunakan python

4. Rubrik Tugas dan Nilai

Kriteria Penilaian: Pemahaman mengenai fondasi Data Pipelines & ETL, paham data engineering

Rubrik Tugas TM021-IMR-TI dan Data-Tugas02			
No Rubrik	1	2	3
Nama Rubrik	Kesesuaian output dengan metode penarikan data	Proses ETL	Visualisasi data

100%	30%	30%	40%
(diisi oleh MNIT nilai final)	(diisi oleh MNIT nilai final)	(diisi oleh MNIT nilai final)	(diisi oleh MNIT nilai final)
75-100	Dua metodenya memiliki output yang sesuai	Men-drop salah satu kolom dan mengintegrasikan beberapa tabel	Visualisasi data lengkap (real time dan batch processing)
50-<75	Salah satu metodenya memiliki metode yang sesuai	Men-drop salah satu kolom namun tidak mengintegrasikan tabel (atau salah satunya)	Visualisasi data yang ditampilkan hanya salah satu (antara real time atau batch processing)
25-<50	Souce codenya ada namun gagal tereksekusi	Souce codenya ada namun gagal mengintegrasikan dan men-drop tabel	Souce codenya ada namun gagal dalam memvisualisasikan data
<25	Tidak ada souce codenya	Tidak ada souce codenya	Tidak ada souce codenya

Bagian B. Kertas Tugas

Tugas ini adalah Data pipeline dari hashtag business, tempat kami mengambil data yaitu dari Twitter dengan menggunakan API Twitter yang disediakan. Data yang telah terintegrasi selanjutnya divisualisasikan menggunakan Power BI.

1). Metode yang digunakan

- ETL (Extract, Transform, Load)

=> Extract

Pertama dilakukan pemanggilan library dengan coding sebagai berikut:

```
import tweepy
import csv
import pandas as pd
from sqlalchemy import create_engine
from sqlalchemy import insert
from sqlalchemy import MetaData
from sqlalchemy import Table
from sqlalchemy.dialects.postgresql import insert

from textblob import TextBlob
from nltk.corpus import stopwords
import re
```

Setelah library dilakukan proses penarikan data dari twitter menggunakan consumer key , consumer secret, dsb., Proses tersebut menggunakan coding berikut:

```
consumer_key = 'a76nBj6z3L9YZudQUAYM5rH9K'
consumer_secret = 'Oxd42VMcm5ZZPoD3S9M97ek2T6pNlxosW04dsmtPQuqPZmrWic'
access_token = '210015757-Gb93f5McEd3TJ5Uld5HI6wCqFJAfSN5evm4Ufzgi'
access_secret = 'J1qvgXGy7Pd1BjXEz8TE90Wq77e0y2madnP4EQQPs82RZ'
tweetsPerQry = 100
maxTweets = 100000
hashtag = "business"
```

=> Transform

Setelah data diambil , perlu dilakukan transformasi data agar data yang diperoleh dapat dengan mudah dibaca/ dianalisis oleh kelompok kami. Proses tersebut menggunakan coding sebagai berikut :

```
#clean tweet text
def clean_text(text):
    ex_list = ['rt', 'http', 'RT']
    exc = '|'.join(ex_list)
    text = re.sub(exc, ' ', text)
    text = text.lower()
    words = text.split()
    stopword_list = stopwords.words('english')
    words = [word for word in words if not word in stopword_list]
    clean_text = ' '.join(words)
    return clean_text

def sentiment_score(text):
    analysis = TextBlob(text)
    if analysis.sentiment.polarity > 0:
        return 1
    elif analysis.sentiment.polarity == 0:
        return 0
    else:
        return -1

authentication = tweepy.OAuthHandler(consumer_key, consumer_secret)
authentication.set_access_token(access_token, access_secret)
api = tweepy.API(authentication, wait_on_rate_limit=True)
maxId = -1
tweetCount = 0
```

=> Data Loading

Data yang telah ditransform, di-load ke file CSV agar bisa divisualisasikan di power BI.

```
csvFile = open(hashtag+".csv", "a+", newline="", encoding="utf-8")
csvWriter = csv.writer(csvFile)
i = []
u = []
ca = []
tt = []
rc = []
fc = []
ms = []
rs = []

while tweetCount < maxTweets:
    if(maxId <= 0):
        newTweets = api.search_tweets(q=hashtag, count=tweetsPerQry, result_type="recent", tweet_mode="extended")
    else:
        newTweets = api.search_tweets(q=hashtag, count=tweetsPerQry, max_id=str(maxId - 1), result_type="recent", tweet_mode="extended")

    if not newTweets:
        print("Done")
        break
```

2). Integrasi Tabel

Setelah melalui proses ETL dan data pipeline, kelompok kami memilih data tabel yang akan diintegrasikan berdasarkan data yang akan ditampilkan pada visualisasi. Pada proses tersebut kami menggabungkan tabel berdasarkan “UserID” sesuai gambar berikut:

Create relationship

Select tables and columns that are related.

tweet

id	create_at	tweet_text_sent	retweet_count	fav_count
1,46605E+23	01/12/2021 21.21.00	change way worked covid? digital transforming busines...	0	0
1,46605E+23	01/12/2021 21.21.00	business software smb ever need, one, easy-to-use pac...	0	0
1,46605E+23	01/12/2021 21.21.00	check new po folio business oppo unities foreign invest...	0	0

customer

id	username	media_source
1,46604E+23	Hannah COMMISSIONS OPEN PLZ I NEED MONEY UE...	Twitter for iPhone
1,46604E+23	Russ	Twitter for iPhone
1,46604E+23		Twitter for iPhone

Cardinality
Many to many (*)
Cross filter direction
Both

☒ Make this relationship active
☐ Assume referential integrity
☐ Apply security filter in both directions

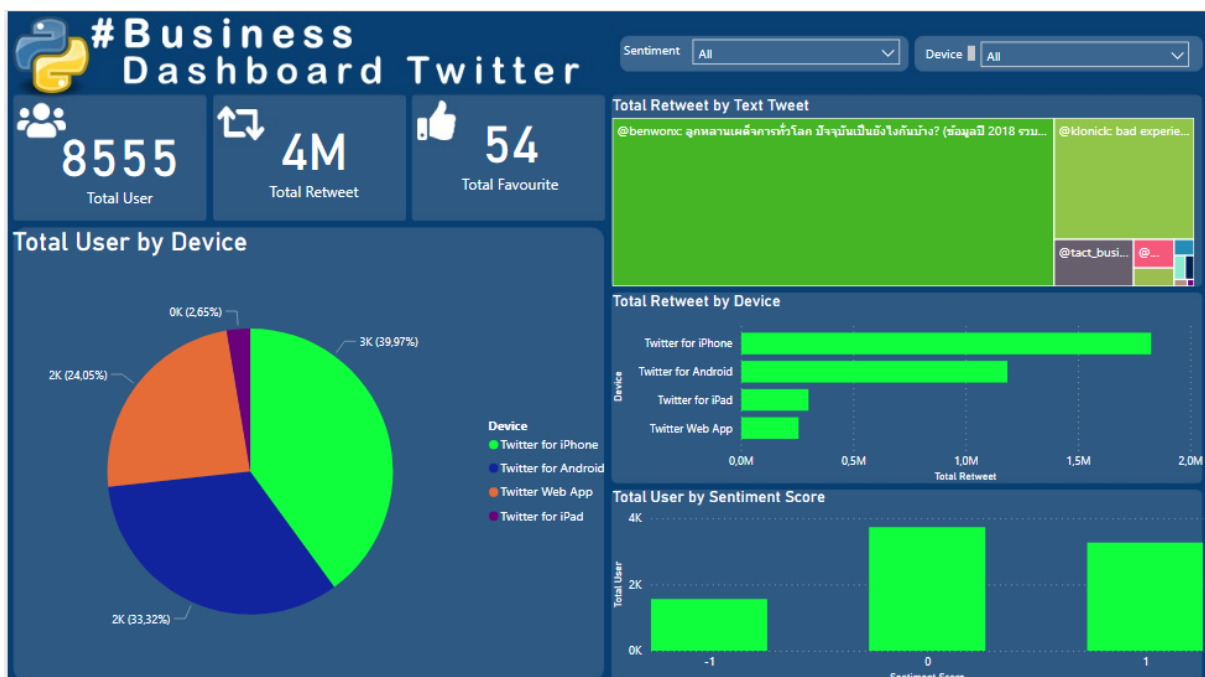
⚠ This relationship has cardinality Many-Many. This should only be used if it is expected that neither column (id and id) contains unique values, and that the significantly different behavior of Many-many relationships is understood.

[Learn more](#)

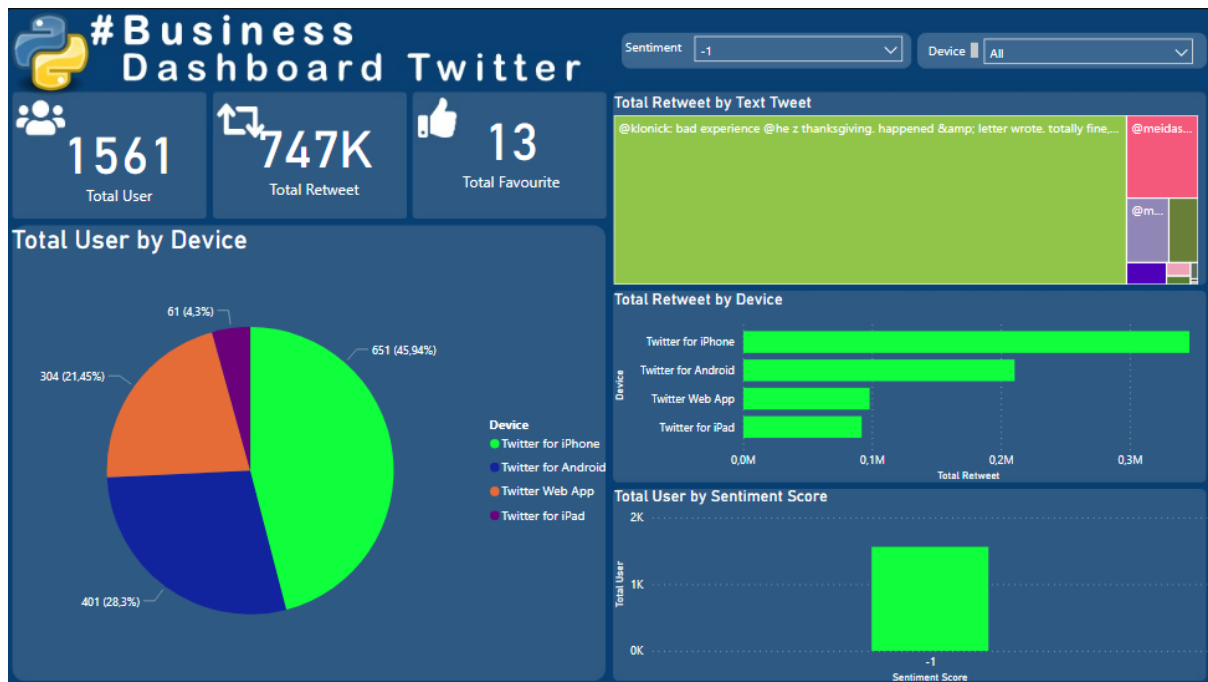
Gambar1. Data Integration

3). Visualisasi Data

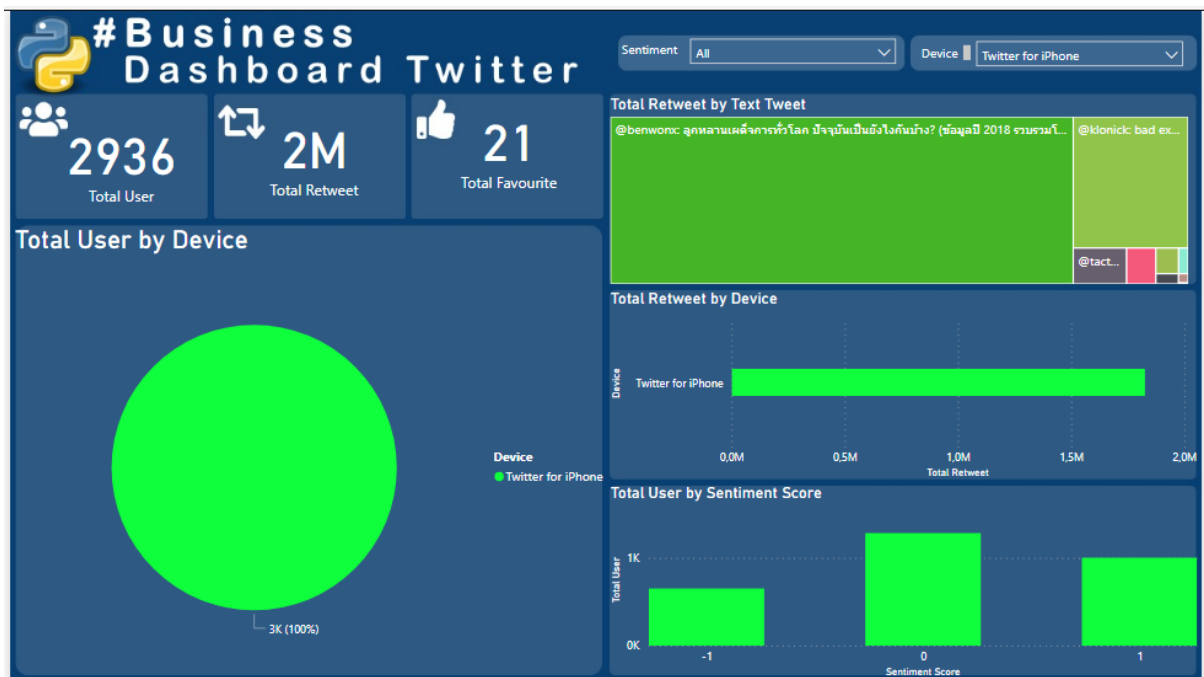
Pada proses visualisasi data kelompok kami menggunakan power BI. Kelompok kami memiliki beberapa fitur untuk menampilkan beberapa pilihan tampilan seperti, berdasarkan device yang digunakan, sentimen masyarakat dan data secara umum. Berikut hasil power BI kami:



Gambar 1. Data #business secara umum



Gambar2. Data #business dengan sentimen negatif



Gambar 3. Data #business pada penggunaan *I-Phone*

Source Code: <https://github.com/Arif-Widagdo/-Business-Dashboard-Twitter.git>.

