

# **MEMBUKTIKAN KESETARAAN GENDER DALAM GAME ONLINE: IMPLEMENTASI DATA ENGINEERING DAN VISUALISASI PENCAPAIAN**

**M Dzikri Mauludil Huda**

**2024**

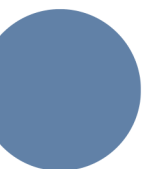
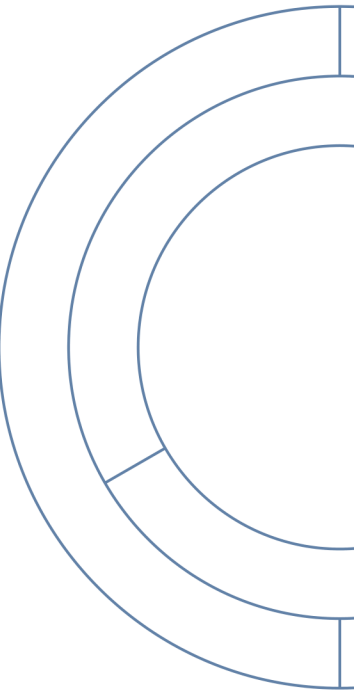


# Latar Belakang



Dalam dunia game online, masih ada stereotip dan stigma yang cukup kuat terhadap pemain perempuan. Banyak konten di media sosial, termasuk TikTok, menunjukkan komentar dan pandangan merendahkan terhadap perempuan yang bermain game, sering kali dianggap sebagai "beban" dalam tim. Namun, apakah benar bahwa perempuan tidak dapat mencapai pencapaian yang sama dengan laki-laki dalam game? Ataukah ini hanya masalah waktu dan kesempatan yang setara?

Hal tersebut menjadi dasar reserach ini dilakukan, dan satu hal menarik lainnya ini dilakukan dengan metoda data engineering untuk mengolah datanya

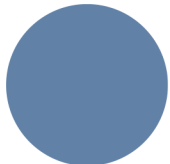


# DATA

df.head(30)

[4]

	PlayerID	Age	Gender	Location	GameGenre	PlayTimeHours	InGamePurchases	GameDifficulty	SessionsPerWeek	AvgSessionDurationMinutes	PlayerLevel	AchievementsUnlocked	EngagementLevel
0	9000	43	Male	Other	Strategy	16.271119	0	Medium	6	108	79	25	Medium
1	9001	29	Female	USA	Strategy	5.525961	0	Medium	5	144	11	10	Medium
2	9002	22	Female	USA	Sports	8.223755	0	Easy	16	142	35	41	High
3	9003	35	Male	USA	Action	5.265351	1	Easy	9	85	57	47	Medium
4	9004	33	Male	Europe	Action	15.531945	0	Medium	2	131	95	37	Medium
5	9005	37	Male	Europe	RPG	20.561855	0	Easy	2	81	74	22	Low
6	9006	25	Male	USA	Action	9.752716	0	Hard	1	50	13	2	Low
7	9007	25	Female	Asia	RPG	4.401729	0	Medium	10	48	27	23	Medium
8	9008	38	Female	Europe	Simulation	18.152733	0	Easy	5	101	23	41	Medium
9	9009	38	Female	Other	Sports	23.942772	0	Easy	13	95	99	36	High
10	9010	17	Male	USA	Strategy	4.829916	0	Hard	8	95	14	12	High
11	9011	36	Female	Asia	Simulation	5.535981	1	Easy	16	124	62	31	High
12	9012	16	Male	USA	Sports	18.776234	1	Easy	9	18	52	32	High
13	9013	38	Female	USA	Strategy	8.701959	0	Easy	0	156	33	47	Low
14	9014	44	Male	USA	Simulation	17.975200	0	Easy	8	41	98	1	Low
15	9015	16	Male	Europe	RPG	7.951511	0	Medium	10	156	58	24	High
16	9016	35	Male	Asia	Strategy	17.887898	0	Easy	16	154	62	4	High
17	9017	47	Male	USA	RPG	17.272113	0	Medium	2	131	13	9	Low
18	9018	26	Female	Asia	Action	19.416079	0	Easy	3	135	77	25	Medium
19	9019	36	Female	Asia	RPG	19.338826	0	Easy	4	56	21	30	Low
20	9020	39	Female	USA	Sports	1.722581	0	Easy	2	177	34	37	Medium
21	9021	41	Female	USA	Sports	16.167690	0	Easy	18	159	36	33	High
22	9022	42	Male	USA	Sports	10.448126	0	Medium	2	120	81	5	Low
23	9023	30	Male	Europe	Strategy	17.959835	0	Easy	14	117	40	19	High
24	9024	29	Male	Europe	Sports	22.564723	0	Medium	12	161	1	22	High
25	9025	17	Female	USA	Sports	10.311460	0	Hard	4	82	1	7	Medium
26	9026	21	Male	USA	Strategy	4.898342	0	Medium	12	118	71	26	High

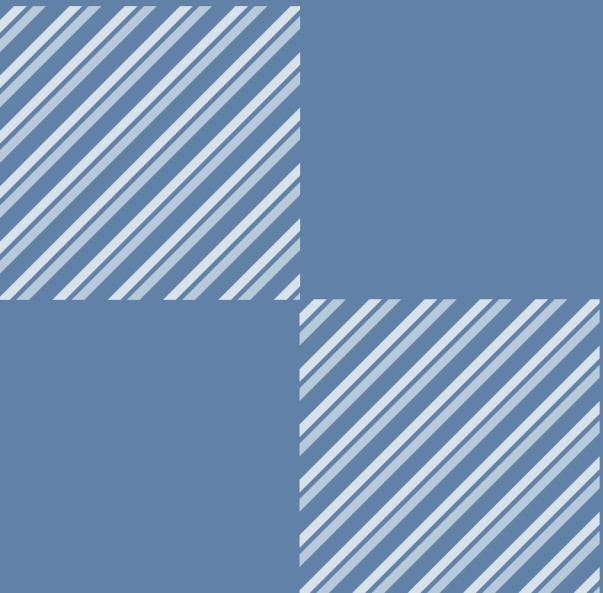
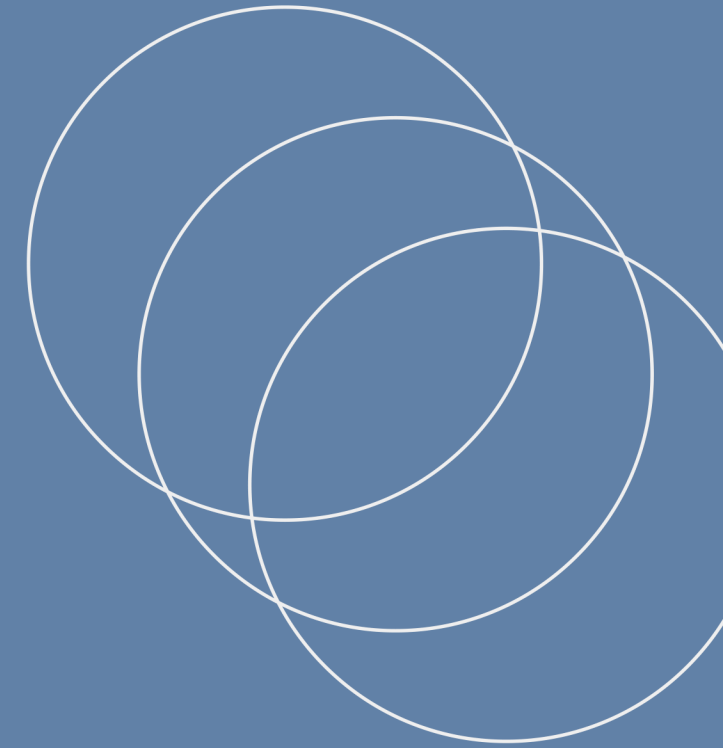




- database
- airflow
- cloud (service account)
- dataproc (hadoop dan spark) as lakehouse
- data gov
- datahub

# Dataset to Mysql

```
You, last week | 1 author (You)
1  from datetime import datetime
2  from airflow import DAG
3  from airflow.operators.python_operator import PythonOperator
4  from airflow.hooks.mysql hook import MySQLHook
5  from airflow.utils.dates import days_ago
6  import pandas as pd
7  from sqlalchemy import create_engine
8
9  def load_data_to_mysql():
10     # conections
11     hook = MySQLHook(mysql_conn_id='mysql_conn2')
12     engine = create_engine(hook.get_uri())
13     #read data
14     df = pd.read_csv('/opt/airflow/csv/online_gaming_behavior_dataset.csv')
15     df.to_sql('tugas', con=engine, if_exists='replace', index=False)
16
17  dag = DAG(
18     'data_transfer_to_mysql',
19     default_args={
20         'owner': 'airflow',
21         'depends_on_past': False,
22         'email_on_failure': False,
23         'email_on_retry': False,
24         'retries': 1,
25     },
26     description='A simple DAG to transfer data to MySQL',
27     schedule_interval='0 17 * * *',
28     start_date=datetime.today()
29 )
30
31  transfer_data = PythonOperator(
32     task_id='load_data_to_mysql',
33     python_callable=load_data_to_mysql,
34     dag=dag,
35 )
36
37  transfer_data
```



# mysql To GCS



```
tags > to_gcs.py > ...
You, 4 days ago | 1 author (You)
1 from airflow import DAG
2 from airflow.operators.python_operator import PythonOperator
3 from airflow.hooks.mysql_hook import MySQLHook
4 from airflow.providers.google.cloud.transfers.local_to_gcs import LocalFilesystemToGCSOperator
5 from airflow.utils.dates import days_ago
6 import pandas as pd
7 import os
8
9 def export_data_from_mysql():
10     hook = MySQLHook(mysql_conn_id='mysql_conn2')
11     engine = hook.get_sqlalchemy_engine()
12     df = pd.read_sql('SELECT * FROM tugas', con=engine)
13     local_path = '/opt/airflow/datasource/datasource.csv'
14     df.to_csv(local_path, index=False)
15     return local_path
16
17 default_args = {
18     'owner': 'airflow',
19     'depends_on_past': False,
20     'email_on_failure': False,
21     'email_on_retry': False,
22     'retries': 1,
23 }
24
25 with DAG(
26     'data_transfer_to_gcs',
27     default_args=default_args,
28     description='DAG untuk mengeksport data dari MySQL dan mengunggah ke GCS',
29     schedule_interval='@once',
30     start_date=days_ago(1),
31 ) as dag:
32
33     export_data_task = PythonOperator(
34         task_id='export_data_from_mysql',
35         python_callable=export_data_from_mysql,
36     )
37
38     upload_to_gcs_task = LocalFilesystemToGCSOperator(
39         task_id='upload_to_gcs',
40         src="{{ task_instance.xcom_pull(task_ids='export_data_from_mysql') }}",
41         dst='testing/real.csv',
42         bucket='testing-de',
43         gcp_conn_id='google_cloud_default', # Gunakan nama koneksi GCS yang benar
44     )
45
46     export_data_task >> upload_to_gcs_task
```

You, last week • add datahub ingestion



# script spark for dataproc dan visualisa



```
proc > spark_job.py > tp main
import matplotlib.pyplot as plt
import pandas as pd
import logging
from google.cloud import storage

def upload_to_gcs(local_file_path, bucket_name, gcs_file_path):
    client = storage.Client()
    bucket = client.get_bucket(bucket_name)
    blob = bucket.blob(gcs_file_path)
    blob.upload_from_filename(local_file_path)

def main():
    logging.basicConfig(level=logging.INFO)
    logger = logging.getLogger('Gaming Data Aggregation')

    spark = SparkSession.builder.appName('Gaming Data Aggregation').getOrCreate()
    try:
        logger.info("Membaca data dari GCS")
        df = spark.read.csv('gs://testing-de/testing/real.csv', header=True, inferSchema=True)
        logger.info("Data berhasil dibaca")

        # Agregasi: Total PlayTimeHours dan AchievementsUnlocked berdasarkan Gender dan GameDifficulty
        logger.info("Melakukan agregasi total PlayTimeHours dan AchievementsUnlocked berdasarkan Gender dan GameDifficulty")
        df_aggregated = df.groupBy('Gender', 'GameDifficulty').agg(
            _sum('PlayTimeHours').alias('TotalPlayTimeHours'),
            _sum('AchievementsUnlocked').alias('TotalAchievementsUnlocked')
        ).toPandas()
        logger.info("Agregasi selesai")
        logger.info("Membuat visualisasi")
        fig, ax = plt.subplots(2, 1, figsize=(12, 12))

        playtime_plot = df_aggregated.pivot(index='GameDifficulty', columns='Gender', values='TotalPlayTimeHours')
        playtime_plot.plot(kind='bar', ax=ax[0], rot=0)
        ax[0].set_title('Total Play Time Hours by Gender and Game Difficulty')
        ax[0].set_ylabel('Total Play Time Hours')

        achievements_plot = df_aggregated.pivot(index='GameDifficulty', columns='Gender', values='TotalAchievementsUnlocked')
        achievements_plot.plot(kind='bar', ax=ax[1], rot=0)
        ax[1].set_title('Total Achievements Unlocked by Gender and Game Difficulty')
        ax[1].set_ylabel('Total Achievements Unlocked')

        plt.tight_layout()
        local_file_path = '/tmp/gaming_data_visualization.png'
        plt.savefig(local_file_path)
        logger.info("Visualisasi selesai dan disimpan sebagai gambar")

        logger.info("Mengunggah gambar ke GCS")
        upload_to_gcs(local_file_path, 'testing-de', 'data/visualizations/gaming_data_visualization.png')
        logger.info("Gambar berhasil diunggah ke GCS")

    except Exception as e:
        You, last week * add datahub ingestion
        logger.error("Error saat menjalankan script", exc_info=True)

    finally:
        spark.stop()
        logger.info("Spark session stopped")

if __name__ == "__main__":
    main()
```

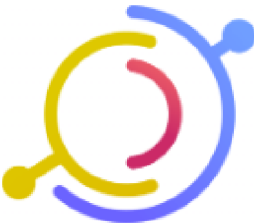
# Ingest data

Home

Analytics

Govern

Ingestion



View all

Search Tables, Dashboards, People, & more...

K

Try searching for


Explore all

general\_log

func

default\_roles


Explore your data



Datasets


331

Platforms



MySQL


336



Google Cloud Storage

2





View all

Search Tables, Dashboards, People, & more...

K

Analytics

Govern

Ingestion

Datasets > Google Cloud Storage > testing-de > testing

Dataset

Google Cloud Storage > testing-de > testing

Share

real.csv

Updated 2 days ago

Schema

Documentation

Lineage

Properties

Queries

Stats

Validation

Incidents

Search in schema...

Field	Description	Tags	Glossary Terms
PlayerID	Number		
Age	Number		
Gender	String		
Location	String		
GameGenre	String		
PlayTimeHours	Number		
InGamePurchases	Number		
GameDifficulty	String		
SessionsPerWeek	Number		
AvgSessionDurationMinutes	Number		

Last synchronized 6 minutes ago

About

No documentation yet. Share your knowledge by adding documentation and links to helpful resources.

Add DocumentationAdd Link

Owners

No owners added yet. Adding owners helps you keep track of who is responsible for this data.

Add Owners

Tags

No tags added yet. Tag entities to help make them more discoverable and call out their most important attributes.

Add Tags

Glossary Terms

No terms added yet. Apply glossary terms to entities to classify their data.

Add Terms

Domain

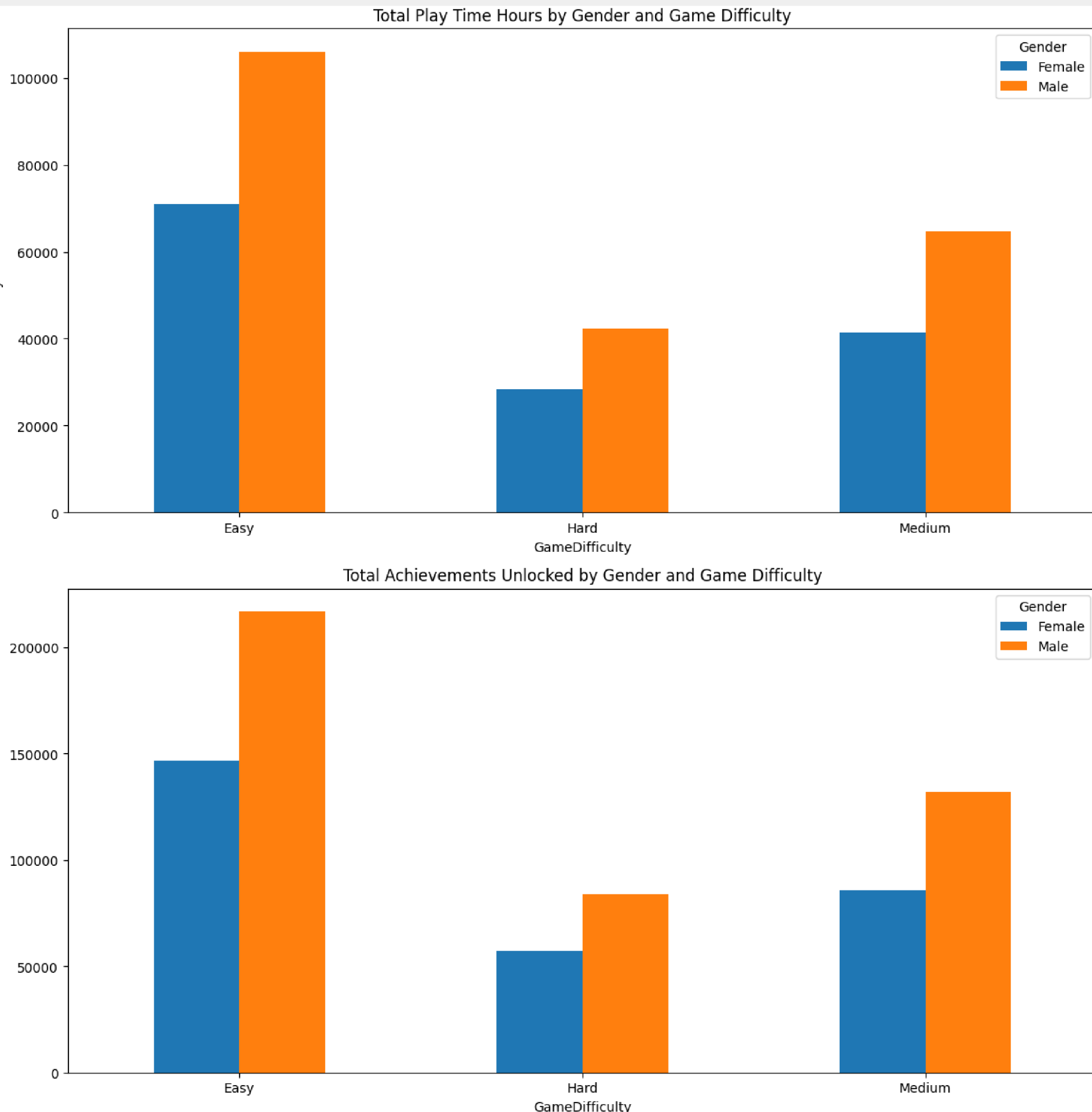
No domain set. Group related entities based on your organizational structure using by adding them to a Domain.

# Hasil viasualisasi

*dari hasil gamabar disamping kita bisa melihat dimana waktu bermain laki laki di tigkat kesulitana manapun itu lebih lama yang mengakibatkan semakain berkembang dan terbiasa dia karena bermain dengan waktu yang lama.*

*dan ini juga yang mengakibatkan achivment yang didapat jelas lebih tingi laki laki karena dengan sering kita bermain maka kia akan sering terbiasa menggunakan nya*

**Berdasarkan analisis, dapat disimpulkan bahwa pencapaian dalam game online tidak bergantung pada gender, melainkan pada waktu bermain yang diinvestasikan. Dengan kesempatan yang setara, pemain perempuan dapat mencapai hasil yang setara dengan pemain laki-laki.**



The background is a solid blue color. It features several geometric elements: a light blue square in the top left; a large white rectangle in the center; three overlapping thin white circles in the top right; a blue rectangle in the bottom right; and two rectangular areas with diagonal blue and white stripes on the bottom left. The text 'Terima Kasih' is centered in the white rectangle, and three small blue squares are positioned below it.

# Terima Kasih

...