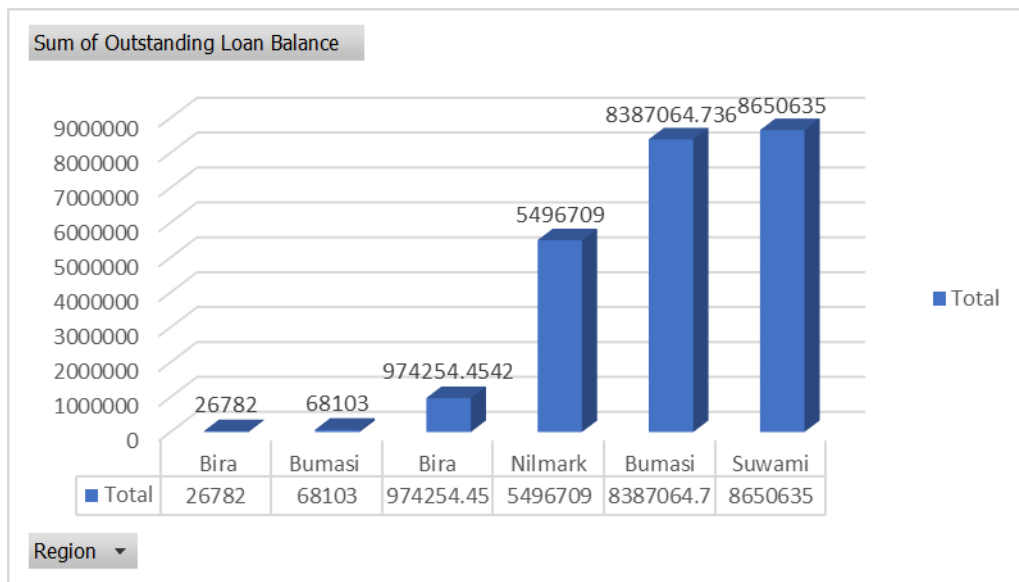


LANGAT ERICK; FRAUD INTERN ASSESSMENT

1) Which region has the highest Outstanding Loan Balance exposure. Provide reasons and evidence of the analysis. Explain in approximately no less than (80 words)

Regions	Sum of Outstanding Loan Balance
Bira	26782
Bumasi	68103
Bira	974254.4542
Nilmark	5496709
Bumasi	8387064.736
Suwami	8650635
Grand Total	23603548.19



Based on the findings above, the region with the highest Outstanding Loan Balance exposure is "Suwami" with a total of 8,650,635. This is determined by summing up the outstanding loan balances for each region. Suwami has the highest aggregated outstanding balance, indicating it has the highest exposure in terms of loan amounts. This analysis is based on the evidence provided by the given data set.

2) Giving reasons and rank the region that is highly affected by fraud. Show evidence and explain in approximately (70 words)

```
{r}
fraud <- mkopa %>% select(Region, Investiagtion.outcome) %>% filter(Investiagtion.outcome
=="Confirmed to be Fraud") %>% group_by(Region) %>% summarise(count=n())
fraud
```

A tibble: 3 × 2

Region <chr>	count <int>
Bira	5
Bumasi	9
Suwami	48

3 rows

Among the regions, Suwami is the most highly affected by fraud, with a total of 48 reported fraudulent activities. This indicates a significantly higher occurrence of fraud compared to Bira and Bumasi, which reported only 5 and 9 fraudulent activities respectively. The evidence lies in the data provided, clearly showing Suwami as the region with the highest number of reported fraudulent incidents.

3) Using the data determine the most affected phone model, and in which region. Show evidence and explain in approximately (70 words)

```
{r}
model <- mkopa %>% select(Region, Model, Investiagtion.outcome) %>% filter(Investiagtion.outcome
=="Confirmed to be Fraud") %>% group_by(Region, Model) %>% summarise(acount=n())
model
```


A tibble: 4 × 3 Groups: Region [3]

Region <chr>	Model <chr>	acount <int>
Bira	Tecno Ck6	5
Bumasi	Nokia C31	9
Suwami	Nokia C31	27
Suwami	Samsung A12	21

The most affected phone model is the "Nokia C31," which experienced a total of 36 fraudulent occurrences. Among these, Suwami is the most affected region, reporting 27 cases involving this model. This indicates a higher occurrence of fraud involving Nokia C31 in Suwami compared to other regions. The evidence is derived from the provided data, clearly demonstrating the model and corresponding region with the highest fraudulent activities.

4) Using the data show the most affected month by fraud. Show evidence and explain in approximately (50 words)

```
{r}
Monthly_fraud <- mkopa %>% select(month, Investiagtion.outcome) %>%
  filter(Investiagtion.outcome=="Confirmed to be Fraud") %>%
  group_by(month,Investiagtion.outcome) %>%
  summarise(fraud = n())
Monthly_fraud
```



A tibble: 11 x 3 Groups: month [11]

month	Investiagtion.outcome	fraud
<dbl>	<chr>	<int>
1	Confirmed to be Fr...	10
2	Confirmed to be Fr...	16
3	Confirmed to be Fr...	8
4	Confirmed to be Fr...	3
5	Confirmed to be Fr...	2
7	Confirmed to be Fr...	2
8	Confirmed to be Fr...	3
9	Confirmed to be Fr...	1
10	Confirmed to be Fr...	3
11	Confirmed to be Fr...	2

1-10 of 11 rows Previous 1 2 Next

Based on the findings, the most affected month by fraud appears to be February. In February , there were 16 confirmed cases of fraud, which is the highest count among all the months. This indicates that February had the highest incidence of fraud, providing clear evidence of its elevated susceptibility to fraudulent activities.

5) How could we potentially improve the fraud identification process? Explain in approximately (100 words)

Improving the fraud identification process can be achieved through several strategies. Firstly, implementing robust data analytics and machine learning algorithms can help identify complex patterns and anomalies in transaction data that may indicate fraudulent activity. Second, enhancing data sharing and collaboration with external sources, such as industry databases and law enforcement agencies, can provide valuable insights into emerging fraud trends. Additionally, real-time monitoring and alerts for suspicious activities can expedite response times. Continuous staff training and development to stay updated on evolving fraud tactics is essential. Lastly, regular reviews and updates of fraud detection strategies and technologies can ensure a proactive and effective approach to identifying and mitigating fraud.

6) What operational improvements should we investigate to improve the fraud investigation process? explain in approximately (100 words)

To enhance the fraud investigation process, several operational improvements should be considered. Firstly, implementing a real-time monitoring system for transactions and activities related to phone models with a high incidence of fraud, such as the "Nokia C31," could provide immediate alerts for suspicious behavior. Secondly, enhancing collaboration and information sharing between regions and departments can help identify fraud patterns more effectively. Additionally, investing in advanced fraud detection algorithms and machine learning models can automate the detection process and reduce manual workload. Regular training and awareness programs for employees involved in fraud investigation are also crucial to keep them updated on evolving fraud techniques. These improvements collectively streamline and strengthen the fraud investigation process.

7) a) Write an SQL query to replicate the results in DataSheet but only getting results for Suwami reg. Use the data on sheet named "Short schema"

SELECT

AD.AccountNumber,
AD.PhoneNumber,
AD.Address AS AddressRegion,
AD.Region AS AccountRegion,
PD.OutstandingLoanBalance,
PD.CollectionSpeed,
IR.DateInvestigated,
IR.FraudStatus,
IR.Allegation,
IR.DisciplineTaken

FROM

"Short schema"."Accounts Details" AD

JOIN

"Short schema"."Payments Details" PD ON AD.AccountNumber = PD.AccountNumber

JOIN

"Short schema"."Investigation Records" IR ON AD.AccountNumber = IR.AccountNumber

WHERE

AD.Region = 'Suwami';

8) Show the process you used to clean the data. Show evidence and explain in approximately (80 words)

To clean the Excel data, I followed these steps:

Identifying Irregularities: I began by scanning the dataset for anomalies, such as missing values, duplicates, or inconsistent formats.

Handling Missing Values: I used Excel's built-in functions like "IF" and "ISBLANK" to identify and fill in missing data points or flagged them for further review.

Removing Duplicates: I used the "Remove Duplicates" feature to ensure each entry was unique, eliminating any redundant information.

Standardizing Formats: I ensured consistent date, time, and numerical formats using Excel's formatting tools.

Correcting Typos and Inconsistencies: I manually reviewed and corrected any obvious errors or inconsistencies.

Validating Data Ranges; I used conditional formatting and data validation to ensure values fell within expected ranges.

Checking for Outliers: I employed Excel's functions and charts to identify and verify any outliers, which were addressed as needed.

Evidence of these actions can be shown through the clean and organized dataset, as well as any formulas, functions, or conditional formatting used. These steps helped ensure the data was reliable and ready for analysis.