1. INTRODUCTION

The purpose of this assessment is to evaluate the following dimensions:
1. Execution quality and accuracy (Ability to deliver a product based on business needs)
2. Problem solving (Ability to understand, learn about and think through new concepts)
3. Data analysis and application (Applying relevant data analysis techniques based on business needs)
4. Self-proficiency (Ability to work independently)
5. Clarity (Communicating deliverables in a clear and concise manner)

1.1 Datasets and Annexes

For this case assessment, there are two datasets and two annexes:
1. Dataset A (HWA-Brovsi-WSA-Claims.csv)
2. Dataset B (HWA-Brovsi-WSA-Claims-Validated..csv)
3. Annex A - Data Dictionary
4. Annex B - WSA Policies

2. CASE CONTEXT

Brovsi is a multinational electronics company that designs and manufactures household appliances such as vacuum cleaners, air purifiers, hand dryers, bladeless fans, heaters, hair dryers, and lights. Its owner, founder and CEO, Eli, had seen the increasing importance of the Asia market to the business and moved the corporate headquarters to Singapore in 2016. His partner, Tedmund, took over as CFO soon after and under his leadership, finance processes have been streamlined significantly with many new information systems platforms introduced to digitise financial data. Recently, he introduced the Welfare Spending Account (WSA) benefits programme in the Singapore head office (Brovsi SG). to allow for employees to manage their welfare benefits according to their needs. WSA claims have three main categories: flexible benefits (FLXI), birthday (BIRT), medical insurance for individuals' scheme (MIIS). Under FLXI, there are eight categories, namely:

1. IT equipment and accessories
2. Childcare Expenses
3. Personal Wellness (Vision Care/Fitness/Exercise)
4. Vacation Expenses
5. Family Insurance Expenses (Spouse, Children, Parents)
6. Personal and Work-life Enrichment
7. Telephone Subscriptions
8. Professional Membership Fees

Having run it as a trial over the last two years, he intends to expand the WSA programme globally. However, they know that there needs to be improvements made to the current system before it can be pushed out globally. The aim of this project is to work out a machine learning plan to analyze data from the 2017 - 2018 WSA claims dataset and in doing so, identify possible outliers and anomalies within so that a predictive tool can be designed to detect possible fraud when the WSA programme goes live for all Brovsi employees around the world. At the same time, if there can be any improvements made to the

current WSA claims system in terms of data collection. From an accounting and finance perspective, to also identify any possible improvements to the current internal controls and policies relating to claims.

2.1 Preliminary Analysis

As part of the data cleansing process, some "rows" of the data has been removed as it has been decided that they will focus primarily only on the approved WSA claims (**Dataset A).** While this approach of "data deletion" to deal with missing data is not ideal, handling missing values in unapproved WSA claims proved to be problematic and unfeasible. Given that the current "cleaned" dataset has no target variables (**Dataset A**), you have been advised to begin the project with **unsupervised learning techniques.** To avoid "domain-bias", you are intentionally not given the full details of the current internal controls put in place within the WSA system. Your preliminary analysis should be able to pick out if the controls have been effective when presenting to the board.

2.2 Risk Analysis

After preliminary data validation, you have been given a validated dataset (**Dataset B).** This dataset is the original dataset (**Dataset A**) with an included target variable ("RISK"). The logic behind how it was determined if a claim was risky is based on internal controls as shown below.

Claim Category	Maximum Allowable Claim Amount ($)
FLXI (1-8)	500.0
MIIS	100.0
BIRT	30.0

Table 1. Allowable Amounts for Claims Categories

Using your preliminary analysis findings, combined with the validated dataset, it is suggested that you attempt **several supervised learning algorithms** and various iterations so that you can make recommendations on how to build a prediction tool to detect possible fraud.

You have also been tasked to look into the claims policy to see if there is a need to tweak it based on your findings, so that the future global WSA claims system will continue to convey the message that staff welfare is the company's priority. There are currently only two WSA claims-related policy and they are as follows:

WSA Policy	Details
WSA Reimbursement	30 to 60 days
WSA Claims Submission	1 to 31 days

Table 2. WSA Claims Policy

3. DELIVERABLES

After you have performed your analysis and developed your models on **Jupyter Notebook,** you will need to present your findings and project plan and findings to the entire senior management, many of whom have no understanding of analytics or machine learning. In your presentation to the board, you should

cover the following

- Basic statistical measures and findings from analysis of dataset variables
- Effectiveness of internal controls based on preliminary analysis
- Machine learning findings based on risk analysis
- Expected outcomes and recommendations

For presentation format, you are strongly encouraged to compile your findings in the **form of slides.** You can also choose to compile your findings in other formats should you deem that it a more intuitive way to present the findings. The key thing is: (1) flow of thinking, and (2) clarity in communications. At the end of the presentation, you may be asked to demonstrate their models on the jupyter notebook and explain some of the components.

Annex A: Data Dictionary

Column Name	Description	Types
EMP_ID	Employee ID	Nil
CLM_SYS	Claim System	**Transport** **Medical** **WSA** **Overtime Expenses** **General Claims** **Small Value Claims** **Marketing Media**
CLM_REF	Claim Reference	Nil
WSA_TYP	Type of WSA Claim	**BIRT**- Birthday FLXI- Flexible Benefits MIIS- Medical Insurance for Individuals' Scheme
FLEXBEN_TYPE	Flexible Benefit Type	**1**- IT Equipment and Accessories 2- Childcare expenses 3- Personal Wellness 4- Vacation Expenses 5- Family Insurance Expenses 6- Personal and Work-life enrichment 7- Telephone Subscription 8- Professional Membership Fees
CLM_DT	Date of Submission of Claim	Nil
CLM_YR	Year of Submission of Claim	
CLM_AMT	Claim Amount ($)	
RCPT_DT	Receipt Date	
RCPT_DAY	Day of Receipt	
DAY_TAG	Day Tag	**Weekday**- Monday, Tuesday, Wednesday,

		Thursday, Friday Weekend- Saturday, Sunday Holiday- Public Holiday
CLM_STAT	Claim Status	**1**- In-Progress 2- Rejected 3- Approved
REIMB_DT	Date of Reimbursement of Claim	Nil
REIMB_YR	Year of Reimbursement of Claim	
EA.DT	External Audit Completion Date	

Annex B: Claim Policies
Allowable Amounts for Claims Categories

Claim Category	**Maximum Allowable Claim Amount ($)**
FLXI (1-8)	500.0
MIIS	100.0
BIRT	30.0

WSA Claims Policy

WSA Policy	**Details**
WSA Reimbursement	30 to 60 days
WSA Claims Submission	1 to 31 days

***Note:** Employees are required to adhere to the above claims policies for their claims to be approved and
reimbursed by their designated or supervising officers. Exceptions are made on a case-by-case basis and must be approved by the Finance Department.