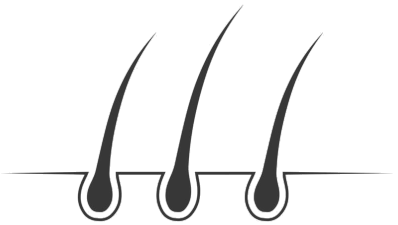


Business Understanding Report

Students Names:

- Alam Bebar - 207415407
- Ryan Thawkho - 305070567

Table of Content

Executive Summary

This report looks at how MyHair, a hair transplant clinic, can use data science to improve its services and patient care. The focus is on combining scalp image analysis with patient data to make the consultation process faster, more accurate, and personalized. It highlights the clinic's goals, current challenges, and how data science could help MyHair deliver better results and stay ahead in the industry.

Report Content

1. Background

1.1. Determining Business Objectives

MyHair, established in 2012 within Skin Clinic and becoming independent in 2015, specializes in hair transplants and aesthetic treatments.[1]

The company aims to enhance patient satisfaction by providing exceptional service and results. Integrating data science can help MyHair better understand factors influencing hair loss and improve treatment outcomes.

1.2. Business Background

- **Available resources:**
  - **Personnel:** A team of experienced doctors, specialists, and support staff.
  - **Material:** Access to patient data, treatment records, and advanced medical equipment.
- **Problems:**
  - Variability in treatment outcomes due to individual patient differences.
  - Limited predictive tools for assessing hair loss progression and treatment efficacy.
- **Goals:**
  - Enhance the accuracy of hair loss assessments.
  - Personalize treatment plans to improve patient satisfaction.
  - Maintain a leading position in the hair transplant industry by adopting innovative technologies.

Task 1 - Determine Organizational Structure

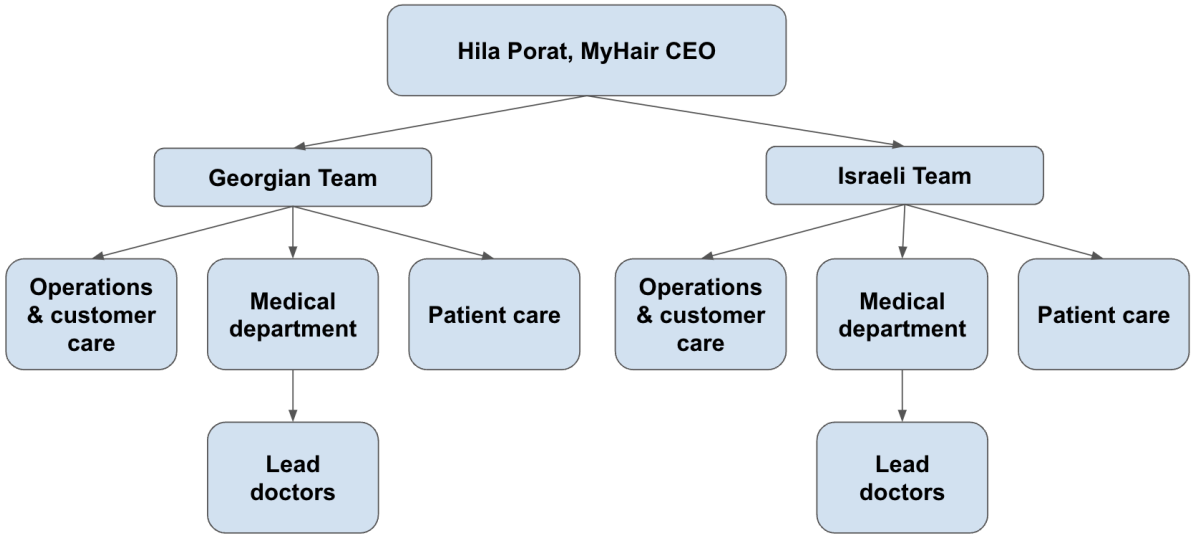
- **Organizational chart:**
  - **CEO:** Hila Polat
  - **Medical department:** Doctors and specialists in hair transplantation.
  - **Patient care:** Support staff managing patient interactions and care.

[1] https://my-hair.co.il/about-my-hair/

- **Operations:** Administrative and logistical support.

- **Key individuals:**
  - **Hila Polat:** CEO, overseeing company operations and strategic direction.
  - **Lead doctors:** Specialists responsible for treatment procedures and patient care, including doctors from Turkey.
- **Affected business units:**
  - **Medical department:** Will utilize data insights to refine treatment protocols.
  - **Patient care:** Can enhance patient interactions based on predictive analytics.
  - **Operations:** May need to adjust workflows to integrate data-driven processes.

The following diagram depicts the organization's chart:



Task 2 - Describe Problem Area

Problem area:

The problem falls under **customer care and operational efficiency**. The focus is on improving the initial consultation process for patients interested in hair transplants.

Problem description:

Currently, MyHair relies on patients submitting photos and forms online or through WhatsApp bots, followed by a manual review process. This approach is time-consuming and resource-intensive, as it requires staff to evaluate images, contact patients, and provide feedback.

The manual process can delay treatment recommendations and doesn't fully utilize technological tools to align patient needs with their current hair loss stage.

Prerequisites and motivation for the project:

- **Prerequisites:**
  - Access to patient-submitted photos and data forms from the clinic would provide more accuracy to our images dataset.
  - Development of machine learning models to automate the evaluation of hair loss stages based on scalp images.
- **Motivation:**
  - Speed up the initial assessment process.
  - Reduce staff workload by automating repetitive tasks.
  - Provide faster, more accurate, and personalized recommendations for patients.
  - Showcase MyHair's use of advanced technology to enhance patient care.

Current use of data science:
MyHair does not currently use data science in its operations. The proposed project introduces a data-driven approach to improve efficiency and patient experience.

Status of the project:
The project is at the planning stage, initiated by students as a proof-of-concept. There's no formal approval needed within MyHair since the project's goal is to explore the potential benefits of automation and data science for the clinic's operations. Presentations to advertise the benefits of data science may be considered after the project demonstrates its feasibility and value.

Task 3 - Describe Current Solution

Current method at MyHair:

- Patients interested in a hair transplant send photos of their scalp and fill out a form through the website or WhatsApp bots.
- The MyHair team manually reviews the images and patient details, then contacts the patient to discuss possible treatments.
- A clinic appointment is scheduled where the hairline and scalp are checked in detail to decide:
  - The number of grafts needed.
  - The type of transplantation method (FUE, DHI, or Hybrid).

- The total price based on the evaluation.

Advantages:

- Personal attention and thorough evaluation during the clinic visit.
- Builds trust with patients through direct interaction.

Disadvantages:

- The process is slow because it relies heavily on manual review and follow-up.
- It requires a lot of staff time and effort to handle submissions and calls.
- The initial steps are not automated, which could save time and resources.

How the project can help:
The project would automate the initial steps, like analyzing patient photos and estimating hair loss stages using machine learning. This would make the process faster, less manual, and more aligned with each patient's current hair loss condition. It would also free up staff time for more important tasks, like in-clinic consultations.

2. Business Objectives and Success Criteria

2.1 Business Objectives

Primary objective:
To enhance MyHair's operational efficiency and improve patient care by developing a system to predict hair loss stages. This system will integrate scalp image analysis and patient data to offer faster, more accurate, and personalized treatment recommendations.

Business questions:

- How can the integration of scalp images and patient data improve the accuracy and speed of initial consultations?
- What insights from these datasets can help MyHair personalize treatments more effectively?
- How can this technology reduce operational costs while improving patient satisfaction?

Other business requirements:

- Ensure **patient data confidentiality** and compliance with healthcare regulations.
- Provide a **user-friendly tool** for MyHair's clinical staff to automate the initial evaluation process.
- Maintain consistency with **MyHair's customer-centric approach** to build trust and loyalty.

Expected benefits:

- **Operational efficiency:** Reduce the time required for the initial consultation process by automating image reviews and data analysis.
- **Patient satisfaction:** Increase patient confidence by providing more precise and data-backed recommendations.
- **Business growth:** Expand customer base by 5% through improved service quality and digital transformation.
- **Cost reduction:** Decrease manual workload, cutting operational costs by 20%.

2.2 Success Criteria

Objective success criteria:

1. Achieve at least **70% accuracy** in predicting hair loss levels using the integrated model.
2. Reduce the time for initial patient evaluations by **30%** by digitizing the initial consultation process.
3. Increase **customer conversion rate** by 5% due to improved efficiency and satisfaction.

Subjective success criteria:

1. Deliver insights that align with MyHair's clinical practices and patient care standards.
2. Ensure that the automated system is well-received by MyHair's clinical team and improves their workflow without adding complexity.

Alignment with stakeholders:

- **Healthcare professionals:** Expect tools that offer actionable insights for improving patient outcomes.
- **Operations team:** Seeks reduced manual effort and faster turnaround times for consultations.

- **Business leaders:** Want measurable improvements in efficiency, patient satisfaction, and financial growth.

3. Assessing the Situation

- **What sort of data are available for analysis?**
    - The project uses two datasets:
        - Over 2000 scalp images from five angles (front, back, left, right, top-down).

The two figures above depict scalp images taken from two angles (left and top-down), which illustrate **Stage 3 of male pattern baldness.**

        - A dataset of 100,000 records of blood component levels with hair loss levels as the target variable. These components include protein, keratine, vitamin, manganese, iron, calcium, body water content, stress level, liver data, and hair fall as the target variable.

# total_protein	# total_keratine	# hair_texture	# vitamin	# stress_level	# hair_fall
0 499	0 499	0 99	0 499	0 99	0 5
312	100	14	249	41	4
52	207	3	425	65	1
170	197	11	140	54	4

An example of four entries from the records dataset displaying data only from 6 columns

- **Do you have the personnel needed to complete the project?**
    - Yes, the project involves a team of data scientists for deep learning and machine learning tasks, as well as domain experts in healthcare to validate the results.
- **What are the biggest risk factors involved?**
    - Integration challenges between the image and numerical datasets.
    - Computational resource requirements for training deep learning models.
    - Ensuring the datasets are representative and of high quality.
- **Do you have a contingency plan for each risk?**

- Yes, contingency plans include using data preprocessing and augmentation techniques to enhance data quality and compatibility, sourcing additional datasets if the existing ones prove insufficient.

4. Inventory of Resources

Task 1 – Hardware Resources

- **What hardware do you need to support?**
  High-performance computing resources, such as GPUs or cloud-based solutions, are required to handle deep learning model training and dataset processing.

Task 2 – Identify Data Sources and Knowledge Stores

- **Which data sources are available for data science? Take note of data types and formats.**
  - Image Dataset: Contains over 2000 images of male scalps from five angles, stored in image file formats (e.g. JPEG or PNG).
  - Numerical Dataset: Contains 100,000 records of blood component levels, likely stored in structured formats such as CSV or databases.
- **How are the data stored? Do you have live access to data warehouses or operational databases?**
  The image dataset is likely stored as individual files, and the numerical dataset is in a downloadable format (such as Kaggle or Roboflow). There is no live access to data warehouses.
- **Do you plan to purchase external data, such as demographic information?**
  There are no current plans to purchase external data, as the existing datasets are enough for this project's scope.
- **Are there any security issues preventing access to required data?**
  No major security issues are anticipated.

Task 3 – Identify Personnel Resources

- **Do you have access to business and data experts?**
  Yes

- **Have you identified database administrators and other support staff that may be needed?**
  Database administrators are not required for this project.

5. Requirements, Assumptions and Constraints

Task 1 - Determine Requirements

Security and legal restrictions: The datasets for this project, sourced from MIT on Kaggle and from Roboflow, have no specific legal or security restrictions mentioned. Both datasets are assumed to follow the platforms' general terms, which typically permit non-commercial research and require proper attribution. Therefore, the datasets can be used in the project without any identified limitations.
Timeline: The team is fully aware of the project's timelines and are committed to meeting the established deadlines. We will make regular check-ins to ensure we are aligned and progressing at the same pace, to ensure the success of this project.
Results deployment: once the model generates predictions, we will anlyze the result using power bi, import the predicted values and generate dashboards that visualize trends, patterns, and key insights from the combined data to aid decision-making.

Task 2 - Clarify Assumptions:

- **Economic factors:** No significant costs are expected since this is a student project, and datasets are freely available.
- **Data quality assumptions:** The existing datasets are assumed to be good enough for a basic model, but preprocessing may be needed to handle inconsistencies.
- **Viewing results:** Since this is a learning project, results will likely be presented simply, such as through accuracy metrics and visualizations, not detailed model interpretations.

Task 3 - Verify Constraints:

- **Passwords for data access:** Not needed; the datasets are public and downloadable.
- **Legal constraints:** No major concerns since public datasets are being used, but respecting data privacy is a priority.

- **Financial constraints:** Covered, as the project relies on free tools and resources like Python libraries and cloud-based platforms like Google Colab.

6. Risks and Contingencies

Risk type	Description	Contingency plan
Scheduling risk	The project may take longer than anticipated due to unforeseen technical challenges or delays in data preprocessing and integration.	Develop a detailed project timeline with buffer periods for potential delays, while prioritizing critical tasks.
Financial risk	Unexpected costs for sourcing additional high-quality data if existing datasets prove insufficient.	Focus on augmenting existing datasets through data preprocessing techniques, such as image augmentation or synthetic data generation
	Exceeding the planned project budget due to unanticipated expenses, such as higher-than-expected computational costs, additional data acquisition, or extended development time	Regularly monitor expenses against the budget with milestone-based cost tracking. Prioritize critical tasks and allocate resources efficiently.
Data risk	The datasets may have poor quality or lack sufficient coverage to generate predictions effectively.	Expand the existing datasets with synthetic data or source additional publicly available datasets if required.

Results risk	The initial results may not meet expectations if it does not achieve a “successful” percentage of accuracy (70%).	Using more advanced models to improve model performance.
	Misinterpretation of results or a lack of actionable insights due to limited domain knowledge	Seeking guidance from domain experts to interpret results and refine the problem scope if necessary.

7. Terminology

Task List

- **List of terms or jargon confusing to team members:**
  - **Multimodal data:** Refers to using two different types of datasets (images and numerical data) together.
  - **Deep learning:** A subset of machine learning used for tasks like analyzing scalp images.
  - **Machine learning:** Techniques used to analyze the numerical dataset (e.g., blood component levels).
  - **Model accuracy:** A performance metric showing how well the model predicts hair loss levels.
  - **Male pattern baldness stages (1-7):** This refers to the **Norwood Scale**, which categorizes hair loss in men from Stage 1 (minimal or no hair loss) to Stage 7 (severe hair loss with only a band of hair around the sides and back).
  - **Hair Recession:** The process where the hairline moves backward, typically at the temples, resulting in an "M" shape. It is one of the early signs of male pattern baldness.
  - **Hair follicle:** The small structure in the scalp that produces and grows hair. Healthy follicles are essential for proper hair growth.

- **Hair loss due to malnutrition:** Hair loss caused by deficiencies in essential nutrients like vitamins (e.g., Vitamin D, B12), minerals (e.g., iron, zinc), or protein. Poor nutrition weakens hair follicles and slows down hair growth.
- **Hair loss due to genetics:** Hereditary hair loss, often referred to as androgenetic alopecia, is caused by genetic factors passed down from either parent, influencing the pattern and severity of hair loss.
- **Hair loss due to hormones (DHT):** Dihydrotestosterone (DHT) is a hormone derived from testosterone. It can shrink hair follicles in genetically predisposed individuals, leading to hair thinning and eventual hair loss.
- **Hair loss preventative methods:** Techniques or treatments aimed at slowing hair loss, such as using **minoxidil**, taking **finasteride**, improving diet, reducing stress, derma rolling, PRP treatments, and avoiding harsh hair treatments.
- **Minoxidil:** A topical medication used to slow hair loss and, in some cases, promote hair regrowth. It is applied directly to the scalp and is commonly used for treating male pattern baldness. It's available over the counter and is most effective when used consistently. Stopping its usage would cause losing any hair regrowth achieved in course of several months.
- **Derma rolling:** This method involves using a handheld device with tiny needles (a derma roller) to create micro-injuries on the scalp. This process stimulates blood circulation, encourages collagen production, and promotes the absorption of topical hair treatments like minoxidil.
- **Finasteride:** An oral medication that helps reduce hair loss by blocking the production of **DHT** (Dihydrotestosterone), the hormone responsible for shrinking hair follicles. It requires a prescription and is effective in treating male pattern baldness, particularly on the crown and mid-scalp areas. Discontinuing finasteride leads to a gradual return of DHT levels to their pre-treatment state. Over several months, hair follicles affected by DHT may shrink again and cause hair loss to resume
- **Hair Transplant Methods (DHI):**
    i. **Direct Hair Implantation (DHI):** A transplant method where hair follicles are extracted and implanted directly into the scalp using a specialized tool, without pre-creating channels.
    ii. **Follicular Unit Extraction (FUE):** A technique where individual hair follicles are extracted from the donor area and implanted into the transplant area. It's minimally invasive and leaves tiny scars.

iii. **Hair Transplant Methods (HYBRID):** A combination of DHI and FUE techniques to utilize the strengths of both methods, offering a tailored approach for better results.

- **Hairline:** The outline of hair growth on the forehead. A natural-looking hairline is a key factor in successful hair restoration procedures.
- **Donor area:** The region, usually at the back or sides of the scalp, where hair follicles are taken for transplantation. This area typically has resistant follicles to hair loss (resistant to DHT).
- **Transplant area:** The part of the scalp where hair follicles are implanted during a hair transplant procedure, usually the balding or thinning regions.

8. Costs and Benefits

Task List

Estimated Costs

- **Data collection:**
  - No additional costs for external data since the clinic can use existing patient-submitted photos and consultation records.
- **Results deployment:**
  - Minimal cost to integrate the system into MyHair's website or clinic workflow.
  - Possible expense for building a simple interface for staff use.
- **Operating costs:**
  - Computational resources for training the models (e.g., cloud-based GPUs) can be managed within a modest budget.
  - Staff training to familiarize the clinic team with the new system.

Benefits

- **Meeting the primary objective:**
  - Faster and more accurate evaluations of hair loss stages, leading to better service delivery and customer satisfaction.
- **Additional insights from data exploration:**
  - Discover new patterns in patient data, such as links between blood components and hair loss severity.

    - Potential to refine treatment protocols based on deeper understanding of contributing factors.
- **Better data understanding:**
    - Improved understanding of patient demographics and needs.
    - Ability to use these insights for marketing strategies and patient retention.

9. Data Science Goals and Success Criteria

9.1. Data Science Goals

Task List

- **Describe the type of data science problem**
    - This project involves classification, as the goal is to predict hair loss levels (severity ratings 1–5) based on both image and numerical datasets.
- **Document technical goals using specific units of time:**
    - Develop and validate a predictive model within a 3-month timeframe.
    - Integrating multimodal data (images and numerical data) to improve the prediction process.
- **Desired outcomes:**
    - Achieve a model accuracy of **at least 70%** in predicting hair loss levels.
    - Demonstrate the feasibility of using the model in clinical or preventive healthcare applications as well as beauty clinics.

9.2. Data Science Success Criteria

Task List

- **Describe the methods for model assessment:**
    - Assess the model based on **accuracy** (target: at least 70%) and performance metrics such as precision and recall to ensure reliable predictions.
- **Define benchmarks for evaluating success:**

- The model must achieve **70% accuracy** or higher on validation data for predicting hair loss severity levels.
- Integration of image and numerical datasets should demonstrate improved prediction quality over using single datasets alone.

- **Subjective measurements:**
  - Evaluate how the model can be used in clinical or preventive applications to provide meaningful insights.
  - The arbiter of success will be the project team, including data scientists, who determine the model's practical utility.
- **Deployment as a success factor:**
  - A simple deployment plan (presenting results in a user-friendly format or prototype app) should be part of the success criteria, **but it will remain basic as we are students.** Deployment is secondary to achieving model performance benchmarks.

10. Project Plan

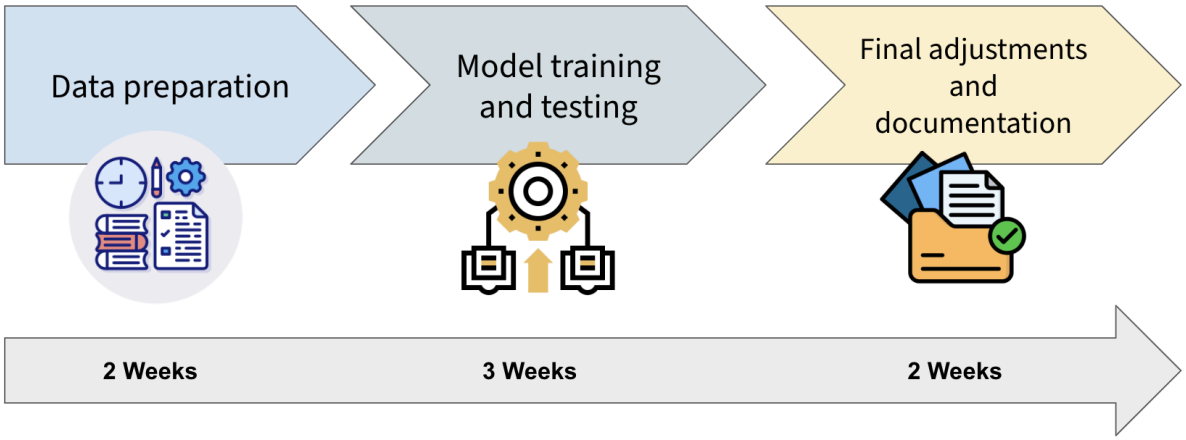
Task List

Have you discussed the project tasks and proposed plan with everyone involved?

- Yes, we've talked about the main tasks, like collecting and preparing data, training the model, testing it, and writing up the results. Everyone on the team knows what they're doing.

Are time estimates included for all phases or tasks?

- Yes, these are the included time phases as follows:



Have you included the effort and resources needed to deploy the results or business solution?

- We'll present the results in a simple format, like graphs or a basic report. No big deployment is needed since this is a student project.

Are decision points and review requests highlighted in the plan?

- Key decisions will be made after the model is first trained and tested. We'll ask for feedback from our professor at those stages.

Have you marked phases where multiple iterations typically occur, such as modelling?

- Training and testing the model will need multiple attempts to improve accuracy and fix issues.

11. Initial Assessment of Tools and Techniques

At this stage, it's still too early to finalize the tools and techniques we'll use for the project. However, we expect to use Python as the primary programming language due to its flexibility and extensive libraries for data science. For the deep learning part, we plan to explore heatmap models

to detect and analyze specific scalp areas from the image dataset. Additionally, we aim to implement an image classification model that can automatically classify scalp images into different stages of hair loss, such as Stage 1 to Stage 7 of male pattern baldness. This will help streamline the process of evaluating patient-submitted images and reduce the need for manual assessments.