



RELATÓRIO EXECUTIVO – PROJETO DE ANÁLISE DE DADOS

1. Título do Projeto

FATORES DE SUCESSO PARA REDUÇÃO DE GORDURA CORPORAL

2. Contexto e Objetivo

Este relatório apresenta os resultados do projeto desenvolvido no Bootcamp de Business Intelligence da WomakersCode - Instituto Localiza, realizado pela squad Nina da Hora. O desafio proposto envolveu a construção de um Pipeline ETL Integrado, seguindo a arquitetura medação, além da criação de um dashboard analítico para investigar um problema definido pelo grupo.

O objetivo do projeto é identificar quais fatores influenciam de forma mais significativa a redução do percentual de gordura corporal, considerando variáveis como dieta, tipo de treino, nível de experiência, frequência semanal de treinos e IMC.

Diante do grande volume de informações divergentes no universo fitness, este estudo visa oferecer uma visão baseada em dados, para apoiar profissionais de saúde, educadores físicos e pessoas interessadas em otimizar sua rotina de treinos e alimentação.

3. Fontes de Dados e Metodologia

Fontes de Dados:

Dataset extraído do Kaggle contendo 20.000 registros sobre estilo de vida, treino, dieta e métricas corporais.

Ferramentas Utilizadas:

Python (Pandas, SQLite)
Power BI
SQLite (banco de dados)

Metodologia – Pipeline ETL (Arquitetura Medalhão):

Bronze: Importação dos dados brutos, sem modificação.

Silver: Padronização, limpeza, tradução de categorias e tratamento de inconsistências.

Gold: Aplicação de regras de negócio, classificação de IMC e construção da tabela final para análise.

Os dados finais foram integrados ao Power BI para criação do dashboard analítico.

4. Principais Insights

A análise dos dados revelou padrões consistentes sobre os fatores que impactam a redução de gordura corporal.

Indicadores gerais:

Média do percentual de gordura: 26,1%
Média de IMC: 24,92
Total de participantes analisados: 20.000

Gênero e experiência

Homens (26,14%) e mulheres (26,07%) possuem médias muito próximas, indicando que o gênero não é um fator determinante.

O nível de experiência (frequência de treinos semanais), mostra um impacto claro:

Avançado: 25,54% masculino e 25,72% feminino.

Intermediário: 26,17% masculino e 26,12% feminino.

Iniciante: 26,17% masculino e 26,12% feminino.

Ou seja, quanto maior a experiência, menor o percentual de gordura, independente do gênero.

Tipo de dieta

As diferenças entre os tipos de dietas são moderadas:

Melhores desempenhos: Vegetariana (25,87%) e Paleolítica (25,95%).

Maior percentual de gordura: Vegana (26,28%).

Os dados estão mais próximos, sugerindo que o tipo de dieta isoladamente não influencia de forma relevante no percentual de gordura, não sendo o fator principal.

Tipo de treino

Os valores também são muito próximos entre si, havendo uma diferença mais relevante quando analisado o tipo de treino por gênero:

Homens: melhor desempenho com cardio (26,08%).

Mulheres: melhor desempenho com força (25,95%).

Os dados indicam que a combinação de treinos é mais relevante que escolher uma única modalidade.

A frequência semanal de treino é o fator de maior impacto

Participantes avançados que treinam de 3 a 5 vezes na semana (25,62%) apresentam menor percentual de gordura e IMC, independente do gênero ou nível de experiência, indicando que consistência e regularidade nos treinos supera a dieta ou modalidade de treino.

Relação entre IMC e Gordura Corporal

O gráfico que relaciona a Classificação do IMC com percentual de gordura mostra que:

Pessoas em sobrepeso/obesidade tem maior gordura corporal.

Quanto maior o IMC, maior o percentual de gordura.

Participantes Avançados chegam a Obesidade Grau I, demais níveis vão até o Grau III

Essa relação também é mostrada no gráfico de correlação, que mostra uma forte correlação positiva.

Em conjunto, os insights revelam que a redução de gordura corporal está mais relacionada à consistência, frequência e experiência, do que ao método isolado (tipo de dieta ou modalidade de treino).

5. Conclusões

A análise dos 20.000 participantes revela que os fatores mais decisivos para redução de gordura corporal estão relacionados ao comportamento, e não exclusivamente ao tipo de dieta ou modalidade de treino. A média geral de gordura permanece em torno de 26%, variando pouco quando analisada isoladamente por dieta ou treino.

As dietas Vegetarianas e Paleolíticas apresentam leve vantagem, mas a diferença é pequena e a dieta Vegana apresenta valor levemente mais alto, mas também com pequena diferença.

O mesmo ocorre com modalidades de treino, em que os treinos de Força (mulheres) e Cardio (homens) apresentam melhores desempenhos, porém todos os tipos analisados permanecem próximos da média geral. Com isso, combinar os treinos de cardio e força é mais relevante que escolher uma única modalidade.

O fator mais relevante é a frequência semanal de treino, especialmente entre 3 e 5 vezes por semana, que se mostrou consistente em todos os grupos. Participantes avançados apresentam melhores resultados, reforçando o impacto da constância ao longo do tempo.

A relação entre IMC e Gordura Corporal mostra forte correlação, indicando que aumentos de IMC acompanham aumentos de gordura, especialmente nas faixas de sobrepeso e obesidade (classificação do IMC). E os participantes com frequência de treino avançada (3 a 5x) não ultrapassam o Grau de Obesidade I, como os demais níveis de experiência.

Em resumo: consistência, progressão e regularidade são os principais fatores que explicam melhores resultados.

6. Recomendações Estratégicas

Com base nos dados analisados, recomenda-se:

1. Priorizar a consistência do treino:

Treinar entre 3 e 5 vezes por semana é o fator mais associado à redução de gordura.

2. Desenvolver treinos ajustados ao nível de experiência:

Iniciantes: aumentar a frequência gradualmente de 1 a 2x, para 3x semanais.

Intermediários: consolidar 3-4x semanais para ampliar resultados.

Avançados: focar em 4-5x semanais para continuar progredindo.

3. Focar em dietas que promovam aderência:

Como as diferenças entre as dietas é pequena, sugere-se priorizar:

- Hábitos consistentes,
- Preferências pessoais,
- Restrições e comorbidades.

4. Integrar treinos de força e cardio:

A combinação apresentou melhores resultados do que uma única modalidade.

5. Apoio especial para IMC elevado:

Indivíduos com sobrepeso/obesidade devem receber:

- Acompanhamento mais próximo,
- Plano alimentar personalizado,
- Progressão gradual de treinos.

6. Campanhas educativas sobre constância:

Reforçar a ideia de que resultados vêm do hábito, da constância ao longo do tempo, e não de estratégias isoladas.

7. Próximos Passos

Para ampliar a utilidade do projeto, recomenda-se:

1. Incorporar novas variáveis no modelo, como:

- Ingestão calórica,
- Horas de sono,
- Tempo de treino,
- Intensidade dos exercícios.

2. Realizar análises mais profundas, considerando:

- Objetivo pessoal,
- Histórico de saúde,
- Evolução mensal.

3. Evoluir o dashboard, adicionando:

- Análises temporais,
- Indicadores individuais e comparativos.

4. Desenvolver uma versão mobile, com recomendações personalizadas.

8. Anexos (opcional)

Exploção dos dados:

```
# carregando o arquivo CSV
df = pd.read_csv("estilo_de_vida.csv")
```

```
# Lendo as primeiras 5 linhas do df
df.head()
```

	Age	Gender	Weight (kg)	Height (m)	Max_BPM	Avg_BPM	Resting_BPM	Session_Duration (hours)	Calories_Burned	Workout_Type
0	34.91	Male	65.27	1.62	188.58	157.65	69.05	1.00	1080.90	Strength
1	23.37	Female	56.41	1.55	179.43	131.75	73.18	1.37	1809.91	HIIT
2	33.20	Female	58.98	1.67	175.04	123.95	54.96	0.91	802.26	Cardio
3	38.69	Female	93.78	1.70	191.21	155.10	50.07	1.10	1450.79	HIIT
4	45.09	Male	52.42	1.88	193.58	152.88	70.84	1.08	1166.40	Strength

5 rows x 11 columns

Extração dos dados e salvamento da camada Bronze:

```
# =====  
# EXTRAÇÃO DE DADOS  
# =====  
  
def coletar_dados(estilo_de_vida_csv, bronze_path = 'estilo_de_vida_bronze.csv'):  
    print(Fore.CYAN + '\nEtapa 1: Extraindo dados...\n')  
    df = pd.read_csv(estilo_de_vida_csv)  
    print('Dados originais:')  
    print(df, '\n')  
    # salvando bronze  
    df.to_csv(bronze_path, index = False)  
    print(Fore.CYAN + f'Etapa 1: Arquivo salvo em: {os.path.abspath(bronze_path)}\n')  
    return df
```

```
linhas = df.shape[0]  
colunas = df.shape[1]  
  
print("Linhas:", linhas)  
print("Colunas:", colunas)  
  
Linhas: 20000  
Colunas: 54
```

Transformação - Carregamento do dataset estilo_de_vida_bronze.csv e escolha de 9 colunas para serem trabalhadas

```
# =====  
# TRANSFORMAÇÃO DOS DADOS  
# =====  
  
def transformar_dados(bronze_path, silver_path: 'estilo_de_vida_silver.csv'):  
    # Escolhendo as colunas a serem trabalhadas  
    print(Fore.CYAN + '\nEtapa 2: Transformando dados...\n')  
    print(Fore.CYAN + '\nEscolhendo as colunas a serem trabalhadas...\n')  
    df_tratado = pd.read_csv(bronze_path)  
    df_tratado = df_tratado[['Gender', 'Age', 'Weight (kg)', 'Fat_Percentage', 'BMI', 'diet_type', 'Workout_Type',  
                             'Workout_Frequency (days/week)', 'Experience_Level']]
```

	Gender	Age	Weight (kg)	Fat_Percentage	BMI	diet_type	Workout_Type	Workout_Frequency (days/week)	Experience_Level
0	Male	34.91	65.27	26.800377	24.87	Vegan	Strength	3.99	2.01
1	Female	23.37	56.41	27.655021	23.48	Vegetarian	HIIT	4.00	2.01
2	Female	33.20	58.98	24.320821	21.15	Paleo	Cardio	2.99	1.02
3	Female	36.69	93.78	32.813572	32.45	Paleo	HIIT	3.99	1.99
4	Male	45.09	82.42	17.307319	14.83	Vegan	Strength	4.00	2.00

Renomear as colunas:

```
# Renomeando as colunas  
print(Fore.CYAN + '\nRenomeando as colunas...\n')  
df_tratado = df_tratado.rename(columns={  
    "Gender" : "genero",  
    "Age" : "idade",  
    "Weight (kg)" : "peso",  
    "Fat_Percentage" : "percentual_de_gordura",  
    "BMI" : "imc",  
    "diet_type" : "tipo_de_dieta",  
    "Workout_Type" : "tipo_de_treino",  
    "Workout_Frequency (days/week)" : "frequencia_de_treino_semanal",  
    "Experience_Level" : "nivel_de_experiencia"  
})  
print(df_tratado.head())
```

Criação das listas e tradução das informações dos dados brutos:

```
# Criando listas  
print(Fore.CYAN + '\nCriando listas...\n')  
df_tratado["genero"].unique().tolist()  
df_tratado["tipo_de_dieta"].unique().tolist()  
df_tratado["tipo_de_treino"].unique().tolist()  
df_tratado["nivel_de_experiencia"].unique().tolist()  
  
# Traduzindo  
print(Fore.CYAN + '\nTraduzindo para português...\n')  
df_tratado["genero"] = df_tratado["genero"].map({  
    "Male": "Masculino",  
    "Female": "Feminino"  
})
```

```
df_tratado["tipo_de_dieta"] = df_tratado["tipo_de_dieta"].map({  
    "Vegan": "Vegana",  
    "Vegetarian": "Vegetariana",  
    "Paleo": "Paleolítica",  
    "keto": "Cetogênica",  
    "low-Carb": "Baixo Carboidrato",  
    "Balanced": "Balanceada"  
})  
df_tratado["tipo_de_treino"] = df_tratado["tipo_de_treino"].map({  
    "Strength": "Força",  
    "HIIT": "HIIT",  
    "Cardio": "Cardio",  
    "Yoga": "Yoga"  
})
```

Arredondamento dos valores para 1 casa decimal, transformação das colunas para número inteiro, substituição dos códigos de experiência por categorias e salvamento da camada Silver:

```
# Arredondamento para 1 Casa Decimal
print(Fore.CYAN + '\nArredondando número para uma casa decimal...\n')
df_tratado["percentual_de_gordura"] = df_tratado["percentual_de_gordura"].round(1)
df_tratado["peso"] = df_tratado["peso"].round(1)

# Transformando colunas para int
print(Fore.CYAN + '\nTransformando colunas para int...\n')
colunas = ['frequencia_de_treino_semanal', 'idade', 'nivel_de_experiencia']
df_tratado[colunas] = df_tratado[colunas].astype(int)

# Substituição dos Códigos de Experiência por Categorias
print(Fore.CYAN + '\nSubstituindo códigos de experiência por categorias...\n')
df_tratado["nivel_de_experiencia"] = df_tratado["nivel_de_experiencia"].map({
    1: "Iniciante",
    2: "Intermediário",
    3: "Avançado"
})
print(df_tratado.head())
# salvando a camada Silver
df_tratado.to_csv(silver_path, index = False)
print(Fore.CYAN + f'Etapa 2: Arquivo salvo em: {os.path.abspath(silver_path)}\n')
```

Início da camada Gold - Classificação do Índice de Massa Corporal (IMC)

```
# =====
# ETAPA GOLD
# =====

print(Fore.CYAN + f'Etapa Gold iniciada\n')
print(Fore.CYAN + f'Criando a função de classificação do índice de massa corporal\n')
return df_tratado

# Função de Classificação do Índice de Massa Corporal
def classificar_imc(imc):
    if imc < 18.5:
        return "Abaixo do Peso"
    elif 18.5 <= imc <= 24.9:
        return "Peso Normal"
    elif 25.0 <= imc <= 29.9:
        return "Sobrepeso"
    elif 30.0 <= imc <= 34.9:
        return "Obesidade Grau I"
    elif 35.0 <= imc <= 39.9:
        return "Obesidade Grau II"
    else:
        return "Obesidade Grau III"
```

Leitura do dataset estilo_de_vida_silver.csv para salvar a camada gold:

```
def gerar_gold(silver_path='estilo_de_vida_silver.csv', gold_path='estilo_de_vida_gold.csv'):
    # lendo o arquivo silver
    df_gold = pd.read_csv(silver_path)

    # Aplica o classificador de forma segura
    df_gold["classificacao_imc"] = df_gold["imc"].apply(classificar_imc)

    # mover para a 6ª posição (índice 5)
    colunas = list(df_gold.columns)
    colunas.insert(5, colunas.pop(colunas.index("classificacao_imc")))
    df_gold = df_gold[colunas]

    # Salva GOLD
    df_gold.to_csv(gold_path, index=False)
    print(Fore.CYAN + f'Etapa Gold concluída e salva em: {os.path.abspath(gold_path)}\n')
    print(df_gold.head())
    return df_gold
```

	genero	idade	peso	percentual_de_gordura	imc	classificacao_imc	tipo_de_dieta	tipo_de_treino	frequencia_de_treino_semanal	nivel_de_experiencia
0	Masculino	34	65.3	26.8	24.87	Peso Normal	Vegana	Força	3	Intermediário
1	Feminino	23	56.4	27.7	23.48	Peso Normal	Vegetariana	HIIT	4	Intermediário
2	Feminino	33	59.0	24.3	21.15	Peso Normal	Paleolítica	Cardio	2	Iniciante
3	Feminino	38	93.8	32.8	32.45	Obesidade Grau I	Paleolítica	HIIT	3	Iniciante
4	Masculino	45	52.4	17.3	14.83	Abaixo do Peso	Vegana	Força	4	Intermediário

Carga - Criação do Banco de Dados:

```
# =====
# CRIAÇÃO DO BANCO DE DADOS
# =====
def carregar_dados(df, df_tratado, df_gold):
    print(Fore.CYAN + '\nEtapa 3: Criação do Banco de Dados...\n')
    print( Fore.YELLOW + '\nCriando um banco SQLite (arquivo estilo_de_vida.db).....')
    # cria o banco
    engine = create_engine('sqlite:///estilo_de_vida.db')
    # salva todas as tabelas no banco criado
    df.to_sql('estilo_de_vida_bronze', con=engine, if_exists='replace', index=False)
    df_tratado.to_sql('estilo_de_vida_silver', con=engine, if_exists='replace', index=False)
    df_gold.to_sql('estilo_de_vida_gold', con=engine, if_exists='replace', index=False)
    print( Fore.YELLOW + '\nBanco de Dados criado.....')
```

Automação do Pipeline de Dados:

```
def pipeline_etl():
    print(Fore.CYAN + '\nETL iniciada...\n')
    estilo_de_vida_csv= 'estilo_de_vida.csv'
    bronze_path = 'estilo_de_vida_bronze.csv'
    silver_path = 'estilo_de_vida_silver.csv'
    gold_path = 'estilo_de_vida_gold.csv'

    # 1) Extração
    df = coletar_dados(estilo_de_vida_csv, bronze_path)
    # 2) Transformação (gera silver)
    df_tratado = transformar_dados(bronze_path, silver_path)
    # 3) GOLD
    df_gold = gerar_gold(silver_path, gold_path)
    # 4) Carregamento
    carregar_dados(df, df_tratado, df_gold)

pipeline_etl()
```

Dashboard com informações gerais:



Percertual de gordura e IMC por gênero:



Gênero Masculino e Feminino - Variação da dieta e treino:

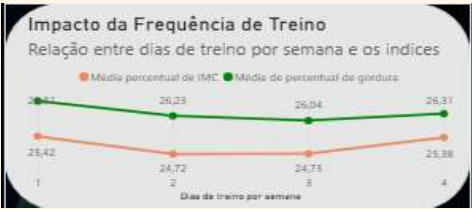


Frequencia semanal de treinos por Nível de experiência - Iniciante, Intermediário e Avançado (geral):



Frequência semanal de treinos, gênero masculino:

Iniciante



Média Percentual de Gordura (%)

26,25

Intermediário



Média Percentual de Gordura (%)

26,17

Avançado



Média Percentual de Gordura (%)

25,54

Frequência semanal de treinos, gênero feminino:

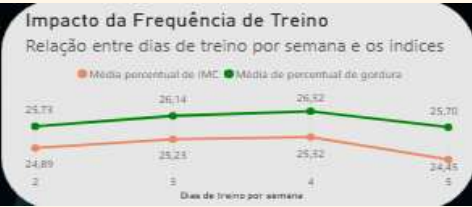
Iniciante



Média Percentual de Gordura (%)

26,11

Intermediário



Média Percentual de Gordura (%)

26,12

Avançado

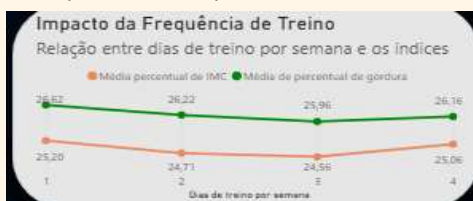


Média Percentual de Gordura (%)

25,72

Frequencia semanal de treinos por Nível de experiência - Iniciante, Intermediário e Avançado (geral):

Iniciante



Média Percentual de Gordura (%)

26,18

Intermediário



Média Percentual de Gordura (%)

26,14

Avançado



Média Percentual de Gordura (%)

25,62

9. Responsável / Equipe

Squad Nina da Hora

Integrantes:

Bruna de Avila Pospiesz

Francielle Cristina da Costa Silva

Gisela Keidel

Ingrid Costa Ferreira

Luana Jaime Tocchio

Pamella Azevedo Costa de Oliveira

Tatiana Varona Villavicencio

Vanelle Rabelo do Nascimento

Vanessa Simão da Costa

25/11/2025