**Divisão das Tarefas**

Como o trabalho é em grupo, a melhor estratégia é dividir as responsabilidades entre os três membros. Podemos seguir o critério das seções do trabalho:

1. **Introdução e Exploração de Dados** (Aluno 1)
	* Definir o problema e contexto (Problem Definition and Setup - 5%)
	* Carregar e entender os dados
	* Fazer análise exploratória (EDA - Exploratory Data Analysis)
	* Identificar insights iniciais (Data Exploration and Findings - 20%)
2. **Preparação e Modelagem** (Aluno 2)
	* Limpeza de dados (Data Cleaning)
	* Engenharia de atributos (Feature Engineering & Feature Selection - 20%)
	* Construção dos modelos de classificação (Classification Models - 20%)
3. **Avaliação, Discussão e Conclusão** (Aluno 3)
	* Avaliação dos modelos (Performance Evaluation - 20%)
	* Comparação dos resultados e insights obtidos
	* Recomendações e conclusões finais (Discussion & Recommendations - 15%)

**Passo a Passo para Concluir o Trabalho**

Aqui está uma estrutura recomendada para seguir:

**Semana 1 - Coleta e Preparação**

1. **Verificar qual é o Problem Set correto para o grupo.**
2. **Baixar o conjunto de dados** correspondente ao problema do grupo.
3. **Familiarizar-se com o notebook template fornecido.**
4. **Dividir oficialmente as tarefas entre os membros do grupo.**
5. **Configurar o ambiente de trabalho (Jupyter Notebook, bibliotecas necessárias).**
6. **Fazer a primeira análise exploratória dos dados (head, info, describe, etc.).**
7. **Identificar problemas nos dados (valores ausentes, outliers, dados inconsistentes).**

**Semana 2 - Análise Exploratória e Engenharia de Atributos**

1. **Fazer análise visual dos dados (histogramas, boxplots, pairplots, correlações).**
2. **Identificar insights iniciais e definir possíveis variáveis importantes.**
3. **Tratar valores ausentes e inconsistentes.**
4. **Criar novas features se necessário.**
5. **Selecionar as melhores variáveis para o modelo.**

**Semana 3 - Construção dos Modelos**

1. **Escolher os modelos de classificação (Logistic Regression, Decision Tree, Random Forest, etc.).**
2. **Dividir os dados em treino e teste.**
3. **Treinar os modelos e otimizar hiperparâmetros.**
4. **Gerar previsões e avaliar os resultados.**

**Semana 4 - Avaliação e Conclusão**

1. **Comparar os resultados dos modelos com métricas adequadas (AUC-ROC, Precision, Recall, Accuracy, etc.).**
2. **Explicar os achados e justificar quais modelos são mais adequados.**
3. **Escrever a conclusão e recomendações com base na análise.**
4. **Finalizar a formatação do notebook e revisar tudo antes da entrega.**
5. **Um membro do grupo faz a submissão final na plataforma Brightspace.**

**Cronograma Sugerido**

| **Semana** | **Tarefa Principal** |
| --- | --- |
| **1** | Baixar os dados, verificar Problem Set e configurar ambiente |
| **1** | Exploração inicial dos dados e identificação de problemas |
| **2** | Análise exploratória e identificação de insights |
| **2** | Preparação dos dados (limpeza, tratamento de outliers, normalização) |
| **3** | Engenharia de atributos e seleção de features |
| **3** | Construção dos modelos de classificação e otimização |
| **4** | Avaliação dos modelos e interpretação dos resultados |
| **4** | Discussão, recomendações e conclusão final |
| **4** | Revisão final e submissão |