Faculdade XPe

RELATÓRIO

PROJETO APLICADO

XP Educação

Relatório do Projeto Aplicado

PLANEJAMENTO DE CULTURAS AGRÍCOLAS NO MUNICÍPIO DE PELICAN TOWN COM BASE EM ANÁLISE DE DADOS E SIMULAÇÃO

Eduardo Silva Coqueiro

Orientador(a): Professor Marcos Prochnow

Janeiro de 2026

EDUARDO SILVA COQUEIRO

XP EDUCAÇÃO

RELATÓRIO DO PROJETO APLICADO

PLANEJAMENTO DE CULTURAS AGRÍCOLAS NO MUNICÍPIO DE PELICAN TOWN COM BASE EM ANÁLISE DE DADOS E SIMULAÇÃO

Relatório de Projeto Aplicado desenvolvido para fins de conclusão do curso de Pós-graduação em Data Science e Machine Learning da XP Educação.

Orientador (a): Professor Marcos Prochnow

Vitória da Conquista - BA

Janeiro de 2026

Sumário

1. CANVAS do Projeto Aplicado

Este Projeto Aplicado tem como objetivo estruturar e analisar um processo de tomada de decisão agrícola em um contexto de pequeno produtor, caracterizado por restrições de capital, tempo e informação. A partir de um cenário inspirado em uma propriedade rural fictícia, o estudo busca compreender como dados relacionados ao cultivo, sazonalidade e qualidade do produto podem apoiar decisões mais assertivas nas fases iniciais da produção. Para isso, são aplicados conceitos de Design Thinking e Data Science, conectando a compreensão do problema à formulação de hipóteses, experimentação analítica e geração de insights práticos para apoio à decisão.

Imagem 1 - Canvas do Projeto Aplicado, apresentando a síntese do desafio, das hipóteses, do objetivo, do backlog e da área de experimentação da solução proposta.

1.1 Desafio

1.1.1 Análise de Contexto

Este Projeto Aplicado parte de um estudo de caso em uma propriedade rural de pequeno porte, denominada Fazenda Stardew, localizada na região do município de Pelican Town. A fazenda é assumida por um novo produtor (o "Dieguinho") que migra de um trabalho urbano para iniciar uma operação agrícola do zero. Embora a narrativa utilize referências do universo de Stardew Valley para facilitar a comunicação e dar identidade ao caso, o problema tratado é real e recorrente na agricultura familiar: planejar o plantio por safra sob restrições severas de tempo, capital e capacidade operacional.

Ao chegar à Fazenda Stardew, o produtor encontra um cenário típico de propriedade em retomada:

- Área produtiva limitada e com necessidade de preparação;
- Infraestrutura inicial básica;
- Dependência de fornecedores locais para sementes e insumos;
- Necessidade de gerar fluxo de caixa rápido para sustentar a operação.

O ambiente socioeconômico local também é coerente com municípios rurais: há um comércio local que atua como principal fornecedor (ex.: Pierre como equivalente a uma casa agropecuária/cooperativa), gestão municipal e calendário de eventos (ex.: Lewis), serviços de apoio e infraestrutura (ex.: Robin/Clint como equivalentes a prestadores de serviço, manutenção e melhorias). Em termos práticos, a fazenda opera com restrições de mercado, logística local e sazonalidade, como ocorre em regiões interioranas.

O desafio central é responder, de forma objetiva e baseada em dados:

Qual estratégia de plantio é mais adequada para cada safra (estação), considerando o ciclo de cultivo e diferentes cenários de qualidade/preço do produto?

Aqui, "estratégia" não significa escolher uma cultura isolada, mas definir um padrão de decisão para cada safra, considerando:

- Janela de tempo para plantio e colheita;
- Custo inicial das sementes (capital de giro);
- Tempo de crescimento (ciclo do cultivo);

- Possibilidade de múltiplas colheitas (quando aplicável);
- Variação de receita conforme a **qualidade do produto** (regular/silver/gold/iridium), interpretada como consequência de manejo/tecnificação.

O objetivo não é "ganhar mais dinheiro" de forma genérica, mas criar um processo replicável de decisão para o produtor iniciante: um guia analítico que indique quais escolhas são mais robustas em cada safra e como a recomendação muda quando a qualidade (manejo) melhora.

O problema é particularmente relevante em propriedades de pequeno porte porque combina decisão econômica com restrições operacionais.

1. Sazonalidade (safras) e janela fixa de produção:

   A fazenda precisa operar em ciclos sazonais. Cada safra tem tempo limitado, e isso impõe:

   - Risco de plantar culturas longas que não "cabem" na janela;
   - Necessidade de priorizar retorno rápido no início da operação;
   - Planejamento de escalonamento e reinvestimento.

2. Capital inicial baixo e dependência de reinvestimento:

   No início, o produtor tem orçamento curto e precisa decidir o que plantar sem comprometer a capacidade de comprar insumos nas próximas etapas. Estratégias que maximizam lucro total, mas demorando muito para retornar, podem ser inviáveis por falta de caixa.

3. Qualidade do produto como determinante de receita:

   A receita agrícola real varia por qualidade, que depende de manejo (ex.: fertilização, irrigação, experiência, tecnologia). No dataset selecionado, essa variação aparece como preços por qualidade (Regular/Silver/Gold/Iridium). Isso permite simular dois perfis realistas:

   - Perfil conservador (baixo investimento em manejo → maior proporção de qualidade regular);

- Perfil otimizado (manejo melhor → maior valor esperado de venda).

4. Decisão multiobjetivo - Uma estratégia “boa” precisa equilibrar:
    - Lucro líquido (rentabilidade);
    - Lucro por dia (velocidade de retorno);
    - ROI (retorno sobre investimento);
    - Risco temporal (plantar e colher dentro da safra);
    - Viabilidade operacional (não exigir decisões complexas demais para um iniciante).

Para viabilizar a análise do desafio proposto, este projeto utiliza um conjunto de dados estruturados que representa informações agronômicas das culturas disponíveis na Fazenda Stardew, incluindo tempo de crescimento, custo de sementes e preços de venda por diferentes níveis de qualidade do produto. Esses dados, obtidos a partir de um ambiente agrícola simulado amplamente documentado, permitem a construção de métricas comparáveis e a simulação de cenários de decisão por safra. A disponibilidade prévia de dados confiáveis e consistentes possibilita que o foco do projeto esteja na modelagem do processo decisório e na avaliação de estratégias, e não na geração ou coleta primária de informações, garantindo reprodutibilidade e clareza analítica.

Origem dos dados:

1. https://www.kaggle.com/datasets/shinomikel/stardew-valley-spring-crop-info?select=spring_crops_info.csv
2. https://www.kaggle.com/datasets/juletopi/stardew-valley-crops-updated

1.1.2 Personas

Persona 1 – Produtor Rural Inicante (Persona Principal):

Nome fictício: Diego “Dieguinho” Martins
Idade: 29 anos
Estado civil: Solteiro
Formação: Ensino superior incompleto (área administrativa/tecnológica)
Origem: Zona urbana
Local de atuação: Fazenda Stardew – região rural de Pelican Town

Contexto pessoal e profissional:

Dieguinho é um jovem adulto que decidiu mudar de vida, deixando um emprego urbano com rotina repetitiva e baixa perspectiva de crescimento para assumir uma propriedade rural de pequeno porte herdada da família. A fazenda encontrava-se parcialmente improdutiva, com infraestrutura básica e necessidade de reestruturação. Sem formação técnica em agronomia, Dieguinho precisa tomar decisões estratégicas de plantio com poucos recursos, alta pressão por retorno financeiro e baixo espaço para erro, especialmente nas primeiras safras.

Persona 2 – Gestor Público Local (Prefeito / Administração Municipal)

Nome fictício: Luiz “Lewis” Almeida
Idade: 52 anos
Estado civil: Casado
Formação: Administração Pública
Origem: Região rural
Local de atuação: Prefeitura de Pelican Town

Contexto pessoal e profissional:

Lewis é o gestor público responsável pela administração municipal de Pelican Town. Seu papel está diretamente ligado à organização do calendário local, eventos sazonais e regras que impactam a atividade econômica da região. Ele busca manter a estabilidade econômica do município, incentivando pequenos produtores rurais a se manterem ativos, uma vez que a agricultura familiar representa parte relevante da economia local. Para Lewis, fazendas sustentáveis significam arrecadação estável e menor êxodo rural.

Persona 3 – Comerciante Local de Insumos Agrícolas

Nome fictício: Pedro “Pierre” Nogueira
Idade: 38 anos
Estado civil: Casado
Formação: Ensino médio completo
Origem: Pelican Town
Local de atuação: Comércio local de sementes e insumos

Contexto pessoal e profissional:

Pierre é proprietário de uma loja local que fornece sementes e insumos agrícolas para pequenos produtores da região. Ele compete diretamente com grandes redes varejistas (representadas pelo grande mercado), que oferecem preços mais baixos, mas menos suporte técnico e menor vínculo comunitário. Pierre depende do sucesso dos produtores locais para manter seu negócio viável e, por isso, tende a incentivar decisões de plantio que gerem retorno rápido e recorrente.

Persona 4 - Representante de Grande Rede Varejista (Concorrência Regional)

Nome fictício: Morris Tod
Idade: 44 anos
Estado civil: Casado
Formação: Administração / Gestão Comercial
Origem: Região metropolitana
Local de atuação: Unidade regional da rede Joja – Pelican Town

Contexto pessoal e profissional:

Morris Tod é o gerente regional responsável pela operação local da **rede varejista Joja**, uma grande empresa de distribuição de insumos e produtos agrícolas que atua em escala regional/nacional. Seu foco está na padronização de processos, redução de custos e aumento de volume de vendas, oferecendo preços competitivos e ampla disponibilidade de produtos.

A presença da Joja em Pelican Town cria um ambiente de **pressão competitiva** sobre o comércio local e, indiretamente, sobre os produtores rurais. Para pequenos produtores como Dieguinho, a rede representa uma alternativa de menor custo imediato, porém com menor suporte técnico, menor vínculo comunitário e decisões menos personalizadas. Essa dinâmica influencia o planejamento agrícola, pois impacta o custo de insumos, a relação de confiança com fornecedores e a sustentabilidade do ecossistema econômico local.

Persona 5 – Prestadora de Serviços e Infraestrutura Rural

Nome fictício: Roberta “Robin” Costa
Idade: 34 anos
Estado civil: Casada
Formação: Técnica em edificações
Origem: Região rural
Local de atuação: Serviços de construção e melhorias rurais

Contexto pessoal e profissional:

Robin atua oferecendo serviços de melhoria estrutural para propriedades rurais, como construções, reformas e adaptações que aumentam a eficiência produtiva. Seu trabalho depende da capacidade financeira dos produtores, sendo normalmente contratado após as primeiras safras bem-sucedidas. Ela representa o elo entre planejamento agrícola e evolução da infraestrutura da fazenda.

Persona 6 – Grande Produtor Rural (Alta Escala e Tecnificação)

Nome fictício: Ricardo “Sr. Ricardo” Andrade
Idade: 47 anos
Estado civil: Casado
Formação: Engenharia Agronômica
Origem: Região rural tradicional
Local de atuação: Propriedade agrícola de médio/grande porte na região de Pelican Town

Contexto pessoal e profissional:

Ricardo é um produtor rural experiente, proprietário de uma fazenda de maior escala e com alto nível de tecnificação. Sua operação conta com melhor infraestrutura, maior capacidade de investimento em insumos, manejo avançado e planejamento de safra mais robusto. Diferentemente de Dieguinho, Ricardo consegue absorver riscos maiores, investir em culturas de ciclo longo e adotar estratégias focadas em **maximização de margem e eficiência operacional**, e não apenas em retorno rápido.

1.1.3 Justificativas

Produtores rurais iniciantes, como o perfil representado pela Fazenda Stardew, enfrentam dificuldades significativas no planejamento agrícola inicial, principalmente relacionadas à escolha de culturas por safra. As decisões de plantio costumam ser baseadas em intuição, recomendações informais ou tentativa e erro, sem o uso sistemático de dados históricos, métricas comparáveis ou simulação de cenários. Esse contexto aumenta o risco financeiro, compromete o fluxo de caixa e pode inviabilizar a continuidade da operação nos primeiros ciclos produtivos.

Além disso, a presença de agentes com diferentes níveis de escala e poder econômico — como grandes produtores tecnificados e redes varejistas de grande porte (ex.: Joja) — amplia a assimetria de informação e competitividade, tornando ainda mais crítica a necessidade de decisões bem fundamentadas para pequenos produtores.

Principais fatores que justificam o desenvolvimento do projeto:

- Dificuldade de produtores iniciantes em definir estratégias de plantio por safra de forma estruturada.
- Alto impacto da sazonalidade e da janela limitada de produção sobre o resultado econômico.
- Risco financeiro elevado associado a decisões de plantio equivocadas no início da operação.
- Dependência de capital de giro e necessidade de retorno rápido para reinvestimento.
- Variação significativa de receita em função da qualidade do produto, relacionada ao nível de manejo.
- Falta de métodos analíticos simples e replicáveis para apoiar decisões agrícolas em pequenas propriedades.
- Assimetria de informação entre pequenos produtores, grandes produtores tecnificados e redes varejistas.
- Potencial de uso de dados estruturados para reduzir incerteza e aumentar a eficiência decisória.

Benefícios futuros esperados:

- Redução do risco financeiro nas primeiras safras por meio de decisões mais embasadas.

- Melhor alocação de recursos (capital, tempo e esforço) ao longo das safras.
- Aumento da previsibilidade do fluxo de caixa, facilitando o planejamento de reinvestimentos.
- Comparação objetiva entre culturas, considerando custo, tempo de crescimento e qualidade.
- Apoio à tomada de decisão estratégica, indo além da escolha de uma cultura isolada.
- Transferibilidade do método para outros produtores de pequeno porte e contextos similares.
- Potencial impacto social, ao contribuir para a sustentabilidade da agricultura familiar e redução do êxodo rural.

Proposta de Valor:

O projeto propõe a construção de um modelo analítico de apoio à decisão agrícola, capaz de recomendar estratégias de plantio por safra a partir de dados estruturados de culturas, considerando restrições reais de tempo, capital e nível de manejo.

A proposta de valor central é:

Transformar decisões agrícolas intuitivas em decisões orientadas por dados, oferecendo ao produtor iniciante um guia claro, replicável e adaptável para maximizar resultados econômicos com menor risco.

Esse valor se materializa ao:

- Traduzir dados agrícolas em métricas comparáveis;
- Simular cenários de qualidade e retorno;
- Gerar recomendações alinhadas à realidade operacional de pequenas propriedades.

1.1.4 Hipóteses

Com base na análise de contexto e na definição das personas envolvidas no desafio, foram identificadas observações centrais relacionadas ao planejamento agrícola em propriedades rurais de pequeno porte. A partir dessas observações, foram formuladas as hipóteses que direcionam o desenvolvimento da solução proposta neste Projeto Aplicado.

Hipótese 1 – Estratégia orientada por dados

Observação: Produtores rurais iniciantes tomam decisões de plantio, majoritariamente, com base em intuição ou recomendações informais.

Hipótese: Se o planejamento de plantio por safra for orientado por dados estruturados, então o desempenho econômico do produtor tende a ser superior em relação a decisões baseadas apenas em intuição.

Grau de risco: Médio

Hipótese 2 – Impacto do tempo de cultivo

Observação: O tempo de crescimento das culturas e a duração da safra nem sempre são considerados de forma conjunta no planejamento agrícola.

Hipótese: Se o tempo de crescimento das culturas for incorporado como variável central na definição da estratégia de plantio, então o retorno econômico por safra tende a ser mais eficiente.

Grau de risco: Baixo

Hipótese 3 – Qualidade do produto e manejo

Observação: A qualidade do produto final impacta diretamente a receita obtida, mas nem sempre é considerada explicitamente nas decisões iniciais de plantio.

Hipótese: Se diferentes cenários de qualidade do produto forem considerados no planejamento agrícola, então as recomendações de estratégia de plantio por safra irão variar de forma significativa.

Grau de risco: Baixo

Hipótese 4 – Adequação ao contexto do produtor iniciante:

Observação: Estratégias utilizadas por produtores de grande escala não são necessariamente viáveis para produtores iniciantes com restrições de capital e infraestrutura.

Hipótese: Se as estratégias de plantio forem ajustadas ao contexto de pequena escala e baixo investimento inicial, então sua aplicabilidade e aderência ao produtor iniciante serão maiores.

Grau de risco: Médio

Observação	Descrição	Criterização			Somatório	Priorização	Grau de risco (Baixo, Médio ou Alto)
		Gravidade (G) 1 a 5	Urgência 1 a 5	Tendência (T) 1 a 6			
Estratégia orientada por dados	Decisões baseadas em dados melhoram o resultado econômico.	5	4	4	13	Alta	Médio
Impacto do tempo de cultivo	O tempo de cultivo influencia a eficiência da safra.	4	4	5	13	Média	Baixo
Qualidade do produto e manejo	A qualidade do produto altera a receita final.	3	4	4	11	Baixa	Baixo
Adequação ao contexto do produtor iniciante	Estratégias precisam se adequar ao produtor iniciante.	4	3	3	10	Média	Média

1.2 Solução

1.2.1 Objetivo SMART

O objetivo deste Projeto Aplicado é desenvolver e validar uma solução analítica de apoio à decisão agrícola, voltada a pequenos produtores, capaz de recomendar estratégias iniciais de plantio a partir da análise de dados de cultivo, sazonalidade, tempo de produção e qualidade do produto. A solução busca reduzir a incerteza nas decisões agrícolas iniciais, transformando dados em recomendações práticas e aplicáveis ao contexto do produtor.

De forma estruturada, o objetivo do projeto é definido segundo os critérios SMART:

- Specific (Específico): desenvolver um modelo analítico que simule e compare diferentes estratégias de plantio, indicando aquelas mais adequadas ao perfil do produtor e ao contexto da safra, considerando variáveis como tipo de cultura, tempo de cultivo e qualidade esperada do produto.
- Measurable (Mensurável): avaliar a solução por meio de métricas comparativas entre cenários simulados, como viabilidade produtiva, tempo estimado de retorno e potencial de qualidade do produto, permitindo verificar a efetividade das recomendações geradas.
- Attainable (Atingível): o objetivo é atingível dentro do prazo e das restrições do projeto, utilizando dados simulados ou históricos, técnicas analíticas compatíveis com o nível do curso e ferramentas acessíveis para pequenos produtores.
- Relevant (Relevante): a solução é relevante por apoiar a tomada de decisão em um contexto recorrente da agricultura de pequeno porte, contribuindo para a redução de riscos, melhor alocação de recursos e aumento da eficiência produtiva.
- Time-based (Temporal): o desenvolvimento, a experimentação e a validação da solução ocorrerão ao longo das três sprints previstas no cronograma do curso, com conclusão e entrega da solução final até o término da Sprint 3, em fevereiro de 2026, conforme planejamento do Projeto Aplicado.

1.2.2 Escopo do Projeto

O escopo deste Projeto Aplicado compreende o desenvolvimento de uma solução analítica baseada em Machine Learning, com foco no apoio à decisão agrícola em cenários iniciais de plantio. A solução será desenvolvida integralmente em Python, utilizando bibliotecas de Data Science e Machine Learning, com experimentação conduzida em ambiente de notebook no Google Colab e dados obtidos a partir de bases públicas disponíveis na plataforma Kaggle.

Para garantir a viabilidade técnica e acadêmica do projeto, foram definidas as seguintes premissas e restrições.

Premissas do Projeto:

- Assume-se que os datasets públicos do Kaggle selecionados representam, de forma aproximada, cenários reais de produção agrícola, sendo adequados para análises exploratórias, modelagem e simulação de estratégias de plantio.
- Considera-se que as variáveis disponíveis nos dados (como tipo de cultura, condições ambientais, tempo de cultivo e indicadores de produtividade ou qualidade) são suficientes para o treinamento e avaliação de modelos de Machine Learning voltados à recomendação de decisões iniciais.
- Parte-se do pressuposto de que os algoritmos de Machine Learning supervisionado, implementados por meio da biblioteca scikit-learn, são apropriados para o escopo do projeto, considerando o nível de complexidade exigido em um Projeto Aplicado de pós-graduação.
- Assume-se a disponibilidade contínua do ambiente Google Colab para desenvolvimento, execução dos experimentos e armazenamento dos notebooks do projeto.

Impactos caso as premissas não se confirmem: a limitação ou inadequação dos dados do Kaggle, bem como a ausência de variáveis relevantes, pode exigir ajustes no escopo analítico, simplificação dos modelos ou redefinição dos critérios de avaliação da solução.

Restrições do Projeto:

- O desenvolvimento da solução está restrito ao prazo definido pelo cronograma do curso, com implementação, experimentação e validação limitadas às três

sprints previstas, devendo a solução final estar concluída ao término da Sprint 3.

- A solução será desenvolvida exclusivamente em ambiente acadêmico, não contemplando deploy em produção, integração com sistemas externos ou uso em operações agrícolas reais.
- O escopo do projeto limita-se ao uso de Python e bibliotecas open source, com destaque para pandas, numpy, matplotlib, seaborn e scikit-learn, não incluindo ferramentas proprietárias ou infraestruturas avançadas de MLOps.
- A modelagem será restrita à fase inicial do ciclo agrícola, não abrangendo etapas posteriores como logística, comercialização ou análise financeira detalhada de longo prazo.

Recursos, Habilidades e Conhecimentos Envolvidos:

- Recursos: Google Colab, notebooks Jupyter, datasets públicos do Kaggle, bibliotecas Python para análise de dados e Machine Learning e Github para organização dos Spirits e versionamento.
- Habilidades e conhecimentos: análise exploratória de dados (EDA), preparação e limpeza de dados, engenharia de atributos, treinamento e avaliação de modelos de Machine Learning supervisionado, interpretação de métricas e comunicação de resultados analíticos.

Validação por métricas:

A validação da solução analítica será baseada em métricas objetivas e comparáveis, alinhadas ao objetivo do projeto. As recomendações geradas pelos modelos serão avaliadas por meio da estimativa de indicadores como lucro esperado, retorno sobre o investimento (ROI) e eficiência temporal (lucro por dia de cultivo), permitindo a comparação entre diferentes estratégias de plantio. No contexto de Machine Learning, a qualidade dos modelos será mensurada por métricas apropriadas a problemas de regressão, como erro médio absoluto (MAE), erro quadrático médio (RMSE) e coeficiente de determinação (R^2), garantindo que as recomendações propostas sejam avaliadas de forma quantitativa, transparente e coerente com o apoio à tomada de decisão agrícola.

1.2.3 Cronograma de Ações Planejadas

O cronograma de ações deste Projeto Aplicado foi estruturado com base na metodologia ágil, considerando a realização de três sprints, conforme o calendário oficial do curso. Cada sprint contempla um conjunto de tarefas necessárias para o desenvolvimento progressivo da solução analítica, desde a preparação dos dados até a validação final do modelo de Machine Learning. O acompanhamento das atividades será realizado por meio de ferramentas de gestão visual, como Trello ou Planner, permitindo o controle do progresso e ajustes ao longo do projeto.

Imagem 2 - Quadro Kanban do projeto, com a organização das tarefas por sprint e status de execução.

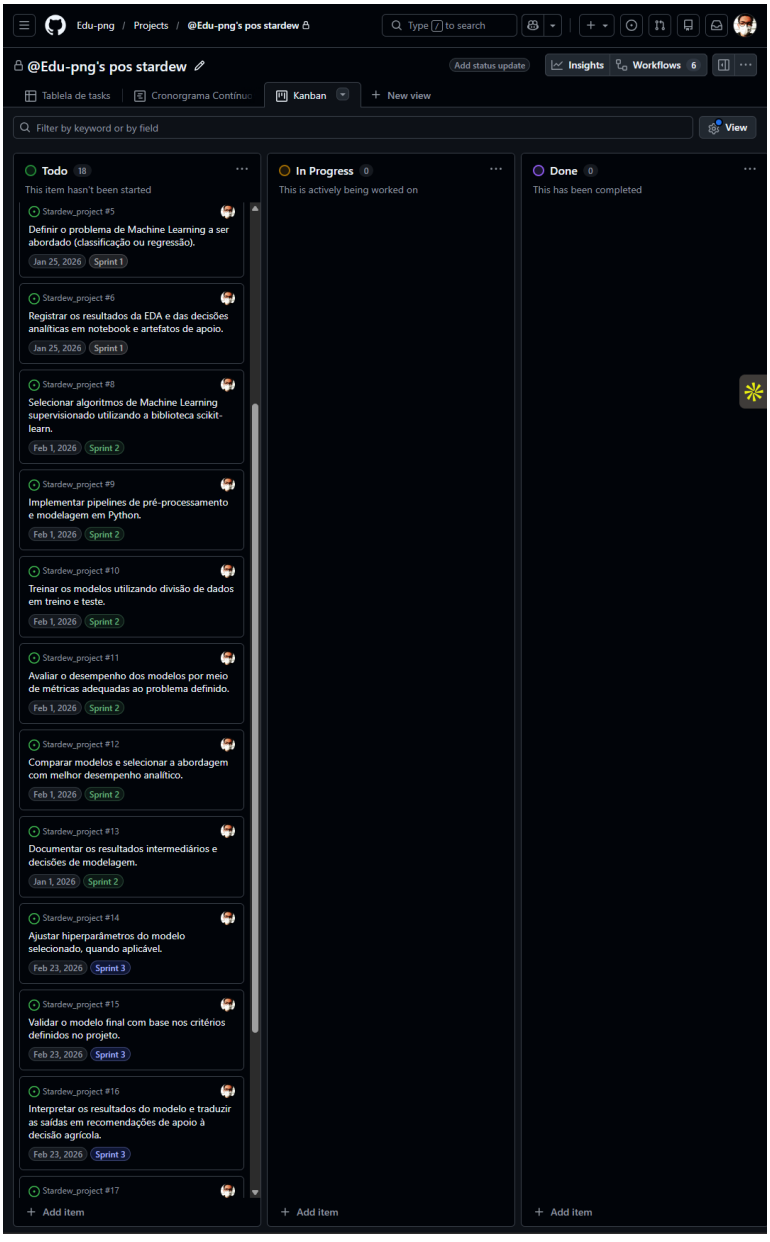
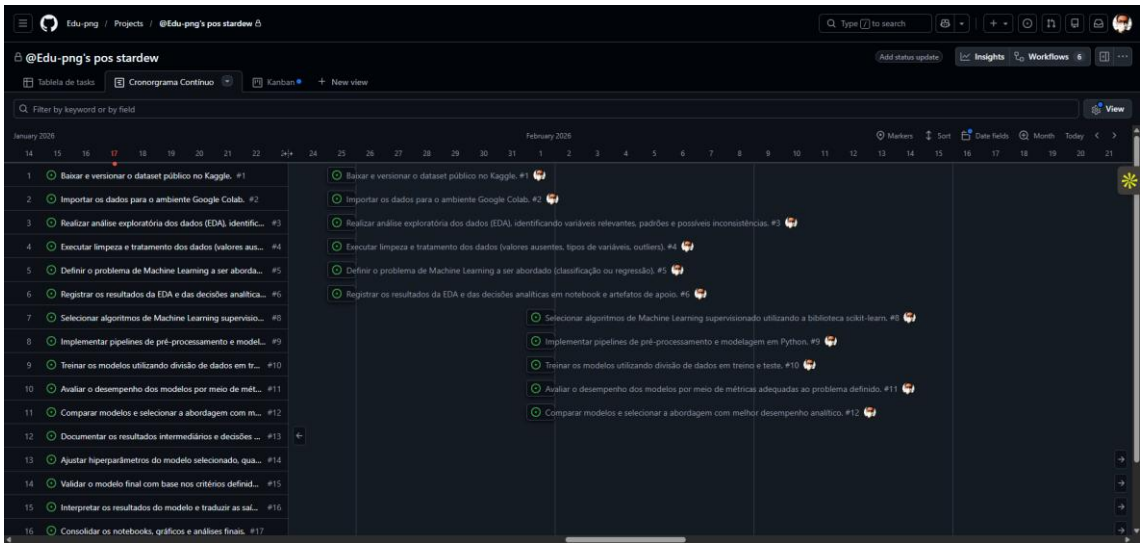


Imagem 3 – Cronograma contínuo do projeto, apresentando a distribuição temporal das tarefas ao longo das sprints.



Sprint 1 – Início do desenvolvimento da solução

Período: até 25/01/2026

Objetivo da Sprint: preparação dos dados e definição da base analítica do projeto.

Tarefas planejadas:

- Baixar e versionar o dataset público no Kaggle.
- Importar os dados para o ambiente Google Colab.
- Realizar análise exploratória dos dados (EDA), identificando variáveis relevantes, padrões e possíveis inconsistências.
- Executar limpeza e tratamento dos dados (valores ausentes, tipos de variáveis, outliers).
- Definir o problema de Machine Learning a ser abordado (classificação ou regressão).
- Registrar os resultados da EDA e das decisões analíticas em notebook e artefatos de apoio.

Entrega da Sprint: base de dados tratada e análise exploratória documentada.

Sprint 2 – Meio do desenvolvimento da solução

Período: 15/01/2026 a 01/02/2026

Objetivo da Sprint: desenvolvimento e treinamento dos modelos de Machine Learning.

Tarefas planejadas:

- Selecionar algoritmos de Machine Learning supervisionado utilizando a biblioteca scikit-learn.
- Implementar pipelines de pré-processamento e modelagem em Python.
- Treinar os modelos utilizando divisão de dados em treino e teste.
- Avaliar o desempenho dos modelos por meio de métricas adequadas ao problema definido.
- Comparar modelos e selecionar a abordagem com melhor desempenho analítico.
- Documentar os resultados intermediários e decisões de modelagem.

Entrega da Sprint: modelos treinados, avaliados e comparados, com escolha do modelo final.

Sprint 3 – Final do desenvolvimento da solução

Período: até 23/02/2026

Objetivo da Sprint: validação da solução e consolidação dos resultados finais.

Tarefas planejadas:

- Ajustar hiperparâmetros do modelo selecionado, quando aplicável.
- Validar o modelo final com base nos critérios definidos no projeto.
- Interpretar os resultados do modelo e traduzir as saídas em recomendações de apoio à decisão agrícola.
- Consolidar os notebooks, gráficos e análises finais.
- Elaborar a documentação final do Projeto Aplicado, incluindo evidências das sprints.
- Iniciar o preparo dos materiais para apresentação à banca avaliadora.

Entrega da Sprint: solução analítica final validada e documentação completa do projeto.

Link do repositório: https://github.com/Edu-png/Stardew_project

2. Área de Experimentação

Ao longo desta seção, a solução é desenvolvida de forma incremental, seguindo a lógica de experimentação orientada por dados. Cada sprint contempla a implementação de requisitos específicos, a execução de experimentos e a análise dos resultados, possibilitando verificar se a abordagem adotada está conduzindo aos objetivos definidos ou se ajustes são necessários (pivotagem). Esse processo garante que a solução evolua de maneira estruturada, transparente e fundamentada em evidências.

Para cada sprint, são apresentados os artefatos que comprovam a execução dos requisitos planejados, incluindo códigos, notebooks analíticos, gráficos, tabelas e registros das decisões tomadas. Além disso, são documentadas as evidências dos resultados alcançados e as lições aprendidas em cada etapa, destacando aspectos que não foram validados, mas que forneceram insights relevantes para o refinamento da solução nas sprints subsequentes.

Essa abordagem experimental reforça o caráter aplicado do projeto, conectando os conceitos de Design Thinking e Data Science à prática de análise, modelagem e simulação de cenários, com o objetivo final de apoiar a tomada de decisão agrícola em um contexto realista de pequeno produtor.

2.1 Sprint 1

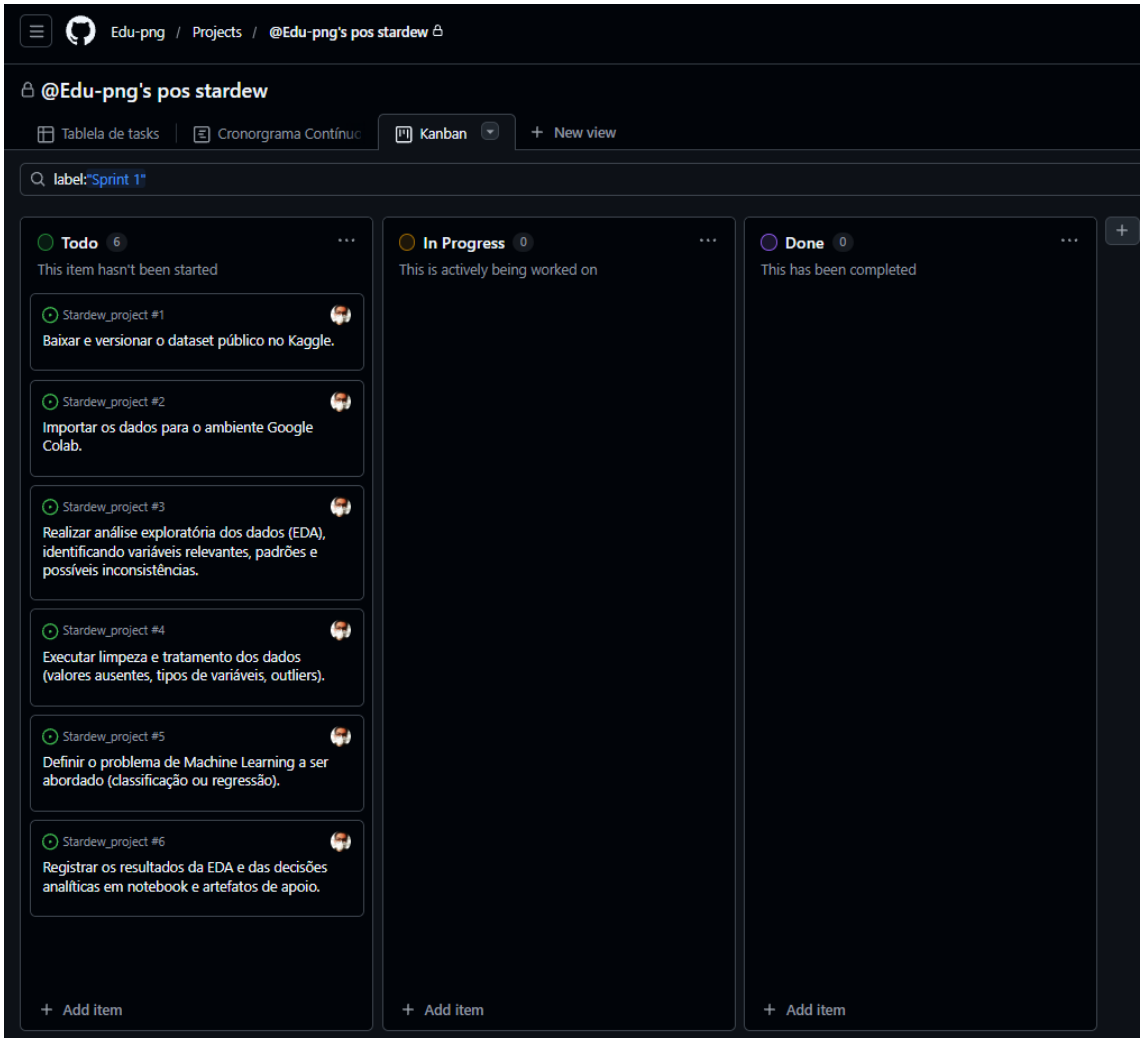
2.1.1 Solução

A Sprint 1 teve como foco a preparação da base analítica do projeto, garantindo que os dados utilizados estivessem adequados para as etapas posteriores de modelagem e simulação de estratégias de plantio. Nesta fase inicial, buscou-se compreender a estrutura dos datasets, identificar variáveis relevantes ao problema e realizar o tratamento necessário para viabilizar análises consistentes e comparáveis.

As atividades desenvolvidas nesta sprint concentraram-se na obtenção dos dados, análise exploratória e organização das informações, estabelecendo as fundações para a construção da solução analítica de apoio à decisão agrícola.

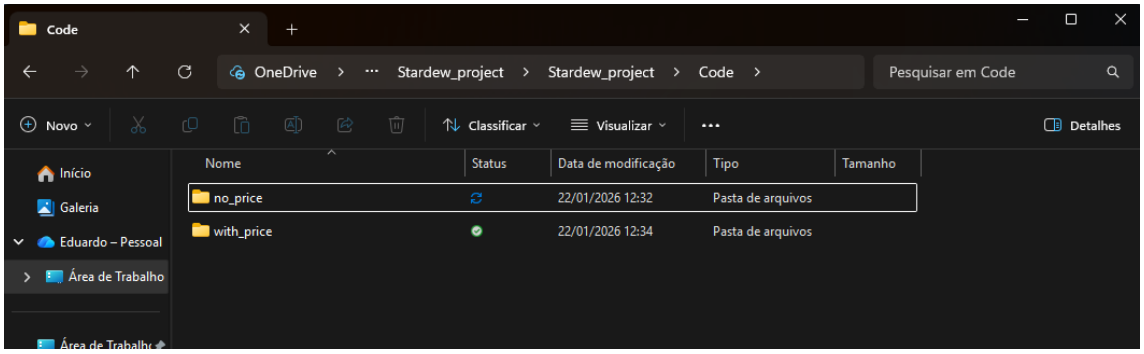
- Evidência da execução de cada requisito:

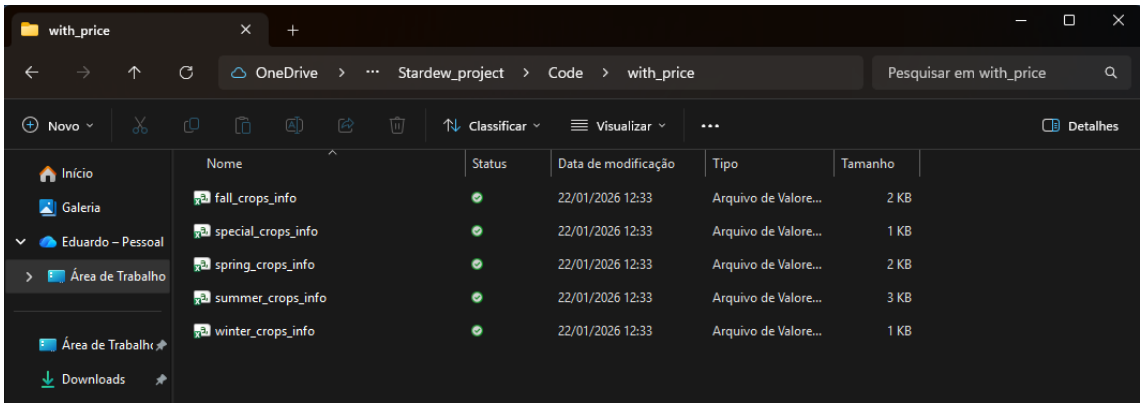
Cards para essa semana (Sprint 1):

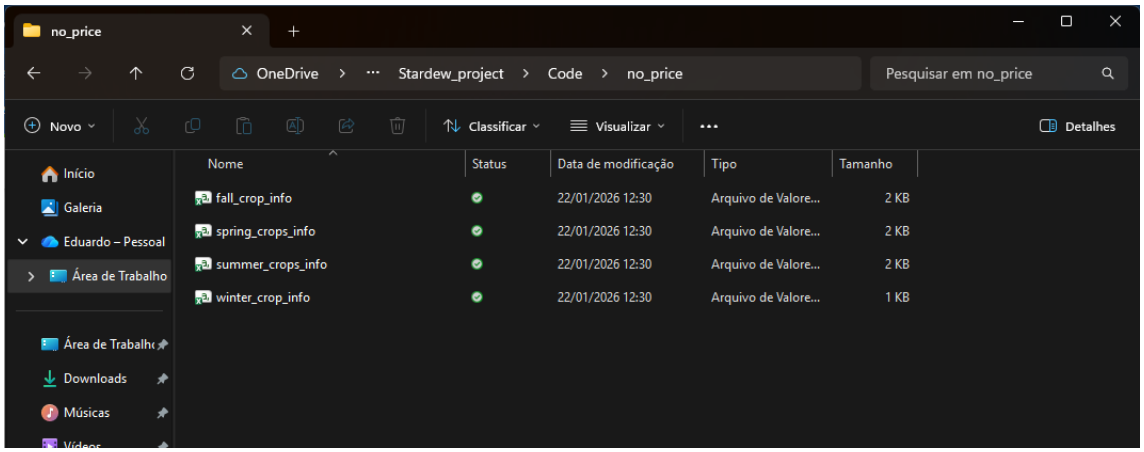


1. Download e versionamento dos datastes públicos da plataforma Kaggle, contendo informações sobre culturas agrícolas, tempo de crescimento, custo de sementes e preços de venda por nível de qualidade:

Imagem 4 - Importando os dados no meu repositório local, tendo dividido os dois datasets em no_price (dataset com informações de venda das culturas mas sem muitos detalhes sobre manuseio) e with_price (mais numérico, com dados de venda e compra).

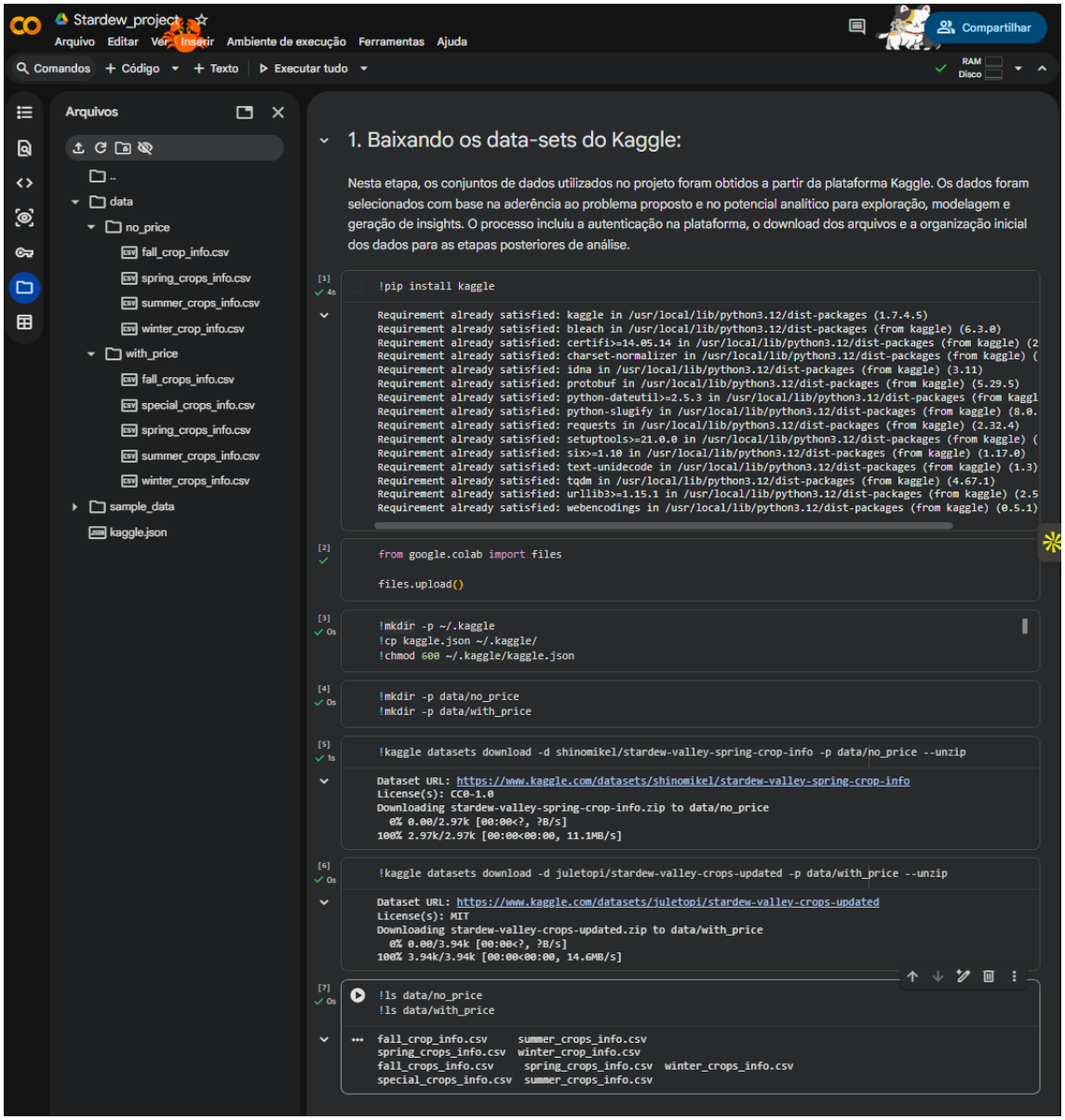






2. Importação dos dados para o ambiente de desenvolvimento no Google Colab, com organização em notebooks.

Imagem 5 - Mesmos imports de data-sets, mas dessa vez no meu ambiente Collab (ocultei a célula que possui minhas credenciais do Kaggle),



3. Realização de Análise Exploratória de Dados (EDA), incluindo inspeção da estrutura dos datasets, tipos de variáveis, identificação de valores ausentes e verificação de consistência das informações.

Imagem 6 – Estatísticas descritivas de um dos data-sets (primavera), que repetem o padrão para as demais com flutuações nos valores numéricos

```
spring_price[["days_to_grow", "seed_price", "sell_price"]].describe()
```

	days_to_grow	seed_price	sell_price
count	14.00	14.00	14.00
mean	7.36	56.07	73.57
std	3.00	29.23	61.53
min	3.00	20.00	15.00
25%	6.00	36.25	31.25
50%	6.50	45.00	45.00
75%	9.50	77.50	102.50
max	13.00	100.00	220.00

Imagem 7 – Resumo das variáveis do dataset de primavera (spring_price), com tipos de dados, classificação analítica e consistência das informações.

```
# Tabela resumo dos tipos de variáveis (dataset de referência)

df = spring_price.copy()

variables_summary = pd.DataFrame({
    "variavel": df.columns,
    "tipo_dado": df.dtypes.values,
    "categoria_variavel": [
        "Numérica" if pd.api.types.is_numeric_dtype(df[col]) else "Categórica"
        for col in df.columns
    ],
    "valores_unicos": [df[col].nunique() for col in df.columns],
    "valores_nulos": [df[col].isnull().sum() for col in df.columns]
})

variables_summary
```

	variavel	tipo_dado	categoria_variavel	valores_unicos	valores_nulos
0	crop_name	object	Categórica	14	0
1	description	object	Categórica	13	0
2	days_to_grow	int64	Numérica	8	0
3	regrowth	int64	Numérica	4	0
4	seed_price	int64	Numérica	9	0
5	sell_price	int64	Numérica	11	0
6	multiple_harvests	object	Categórica	2	0
7	edible	object	Categórica	2	0
8	season	object	Categórica	1	0
9	profit	int64	Numérica	11	0

Próximas etapas: Gerar código com variables_summary | New interactive sheet

4. Limpeza e tratamento dos dados, com padronização de nomes de colunas, ajuste de tipos de dados e exclusão ou correção de registros inconsistentes.

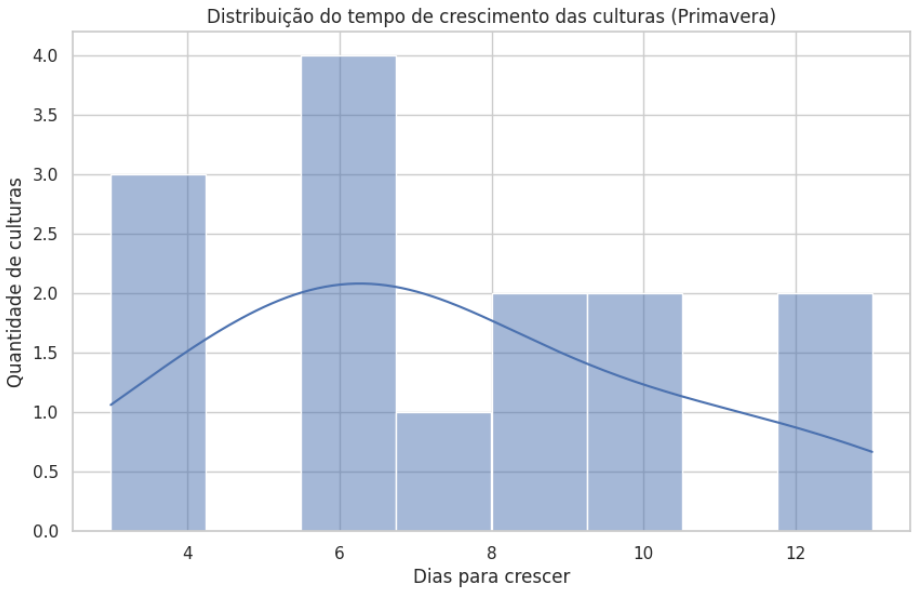
Imagem 8 – Investigação quanto a presença de dados ausentes/faltantes. Não havia essa condição nos data-sets investigados!

```
Dataset: spring_no_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
multiple_harvests    0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: summer_no_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
multiple_harvests    0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: fall_no_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
multiple_harvests    0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: winter_no_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
multiple_harvests    0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: spring_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
regrowth             0
seed_price           0
sell_price           0
multiple_harvests    0
edible               0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: summer_price
Valores nulos por coluna:
crop_name            0
description          0
```

```
--------------------------------------------------
Dataset: summer_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
regrowth             0
seed_price           0
sell_price           0
multiple_harvests    0
edible               0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: fall_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
regrowth             0
seed_price           0
sell_price           0
multiple_harvests    0
edible               0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: winter_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
regrowth             0
seed_price           0
sell_price           0
multiple_harvests    0
edible               0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
Dataset: special_price
Valores nulos por coluna:
crop_name            0
description          0
days_to_grow         0
regrowth             0
seed_price           0
sell_price           0
multiple_harvests    0
edible               0
season               0
dtype: int64
Registros duplicados: 0
--------------------------------------------------
```

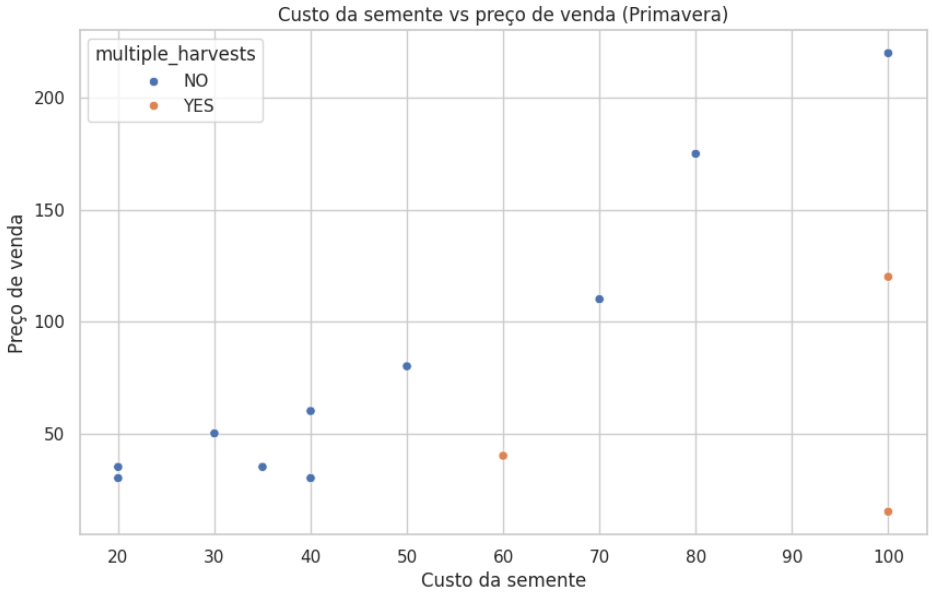
5. Registro das decisões analíticas e dos principais achados da EDA por meio de tabelas, gráficos exploratórios e comentários nos notebooks.

Imagem 9 – Distribuição do tempo de crescimento das culturas da primavera, em dias.



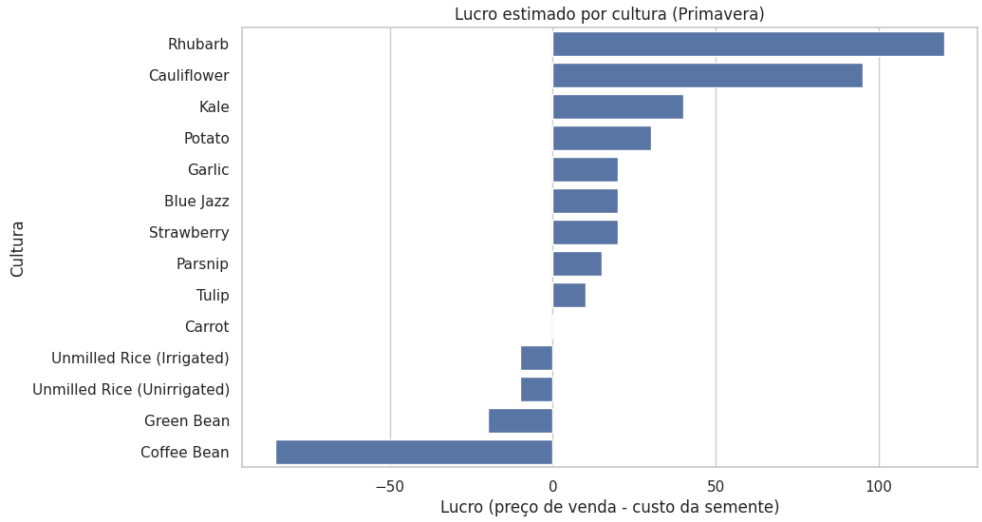
Observa-se concentração de culturas com ciclos mais curtos, indicando predominância de opções com retorno mais rápido, ao mesmo tempo em que culturas de ciclo mais longo podem representar estratégias de maior risco ou planejamento de médio prazo, deve-se levar em conta em etapa posterior que existem culturas que podem ser colhidas mais de uma vez e, portanto, o retorno é maior.

Imagem 10 – Custo da semente vs preço de venda (Primavera)



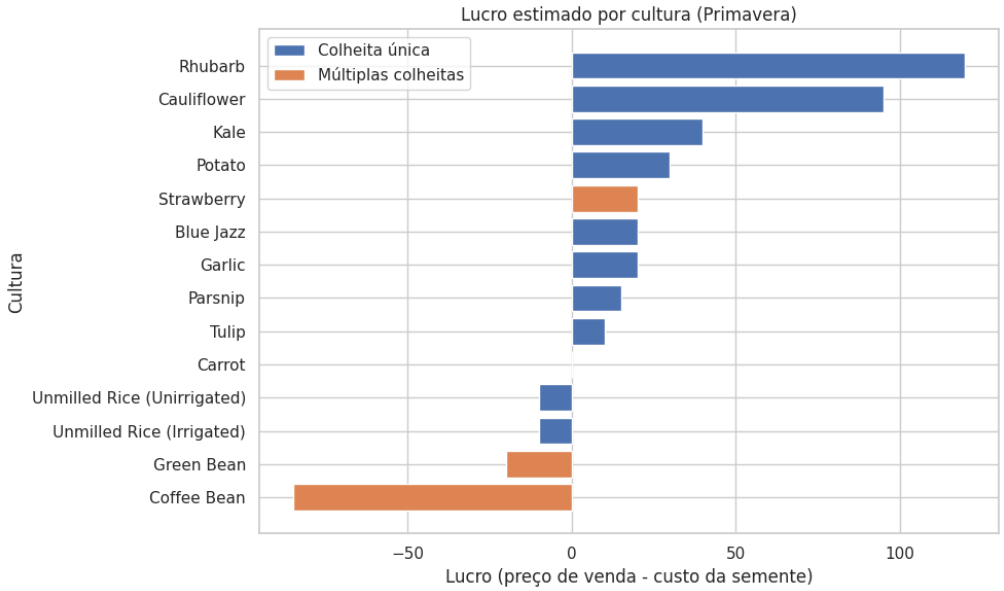
Relação entre o custo da semente e o preço de venda das culturas da primavera, com diferenciação entre culturas de colheita única e colheita múltipla. O gráfico evidencia que maiores custos iniciais nem sempre resultam em maiores retornos unitários, especialmente no caso de culturas com colheita múltipla, cujo retorno ocorre de forma distribuída ao longo do tempo.

Imagem 11 – Lucro estimado por cultura (Primavera)



Lucro estimado por cultura na primavera, calculado como a diferença entre o preço de venda e o custo da semente. O ranking destaca culturas com maior potencial econômico em uma colheita unitária, ao mesmo tempo em que evidencia culturas que apresentam prejuízo inicial, indicando a necessidade de métricas mais adequadas para avaliar culturas de ciclo contínuo ou com rebrote (caso do Café – Coffee Bean por exemplo).

Imagem 12 – Lucro estimado por cultura (Primavera) com separação de cores entre colheita única x múltipla colheitas.



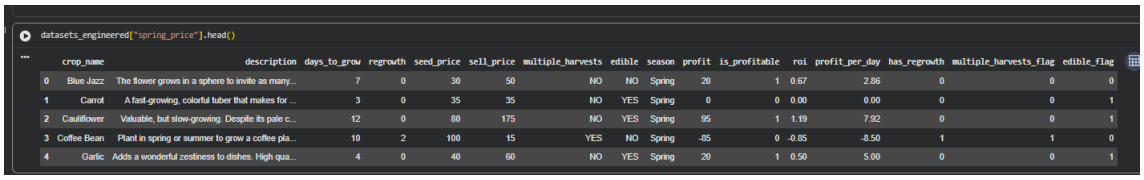
Em continuidade à análise anterior, o gráfico evidencia que culturas de colheita única concentram os maiores lucros unitários no curto prazo. Em contraste, culturas com múltiplas colheitas ou rebrote tendem a apresentar lucros iniciais menores ou negativos, não por inviabilidade econômica, mas por limitações da métrica de lucro unitário em capturar adequadamente ciclos produtivos contínuos. Esse resultado reforça a necessidade de indicadores complementares, como lucro acumulado ao longo do tempo e retorno por dia de cultivo, especialmente para culturas como o Café (Coffee Bean).

7. Etapa adicional: EDA -> Feature Engineering:

A partir dos insights obtidos na EDA, foi identificada a necessidade de incluir uma etapa de Engenharia de Dados. Essa etapa foi adicionada como um novo card no Sprint 1, visando a criação de variáveis estratégicas que melhor representassem os critérios de decisão agrícola.

Imagem 13 – Implementação da etapa de Engenharia de Dados, com criação de variáveis derivadas e aplicação padronizada do processo de feature engineering para todos os datasets com informações de preço.

Imagem 13 – Dataset de primavera após a aplicação da engenharia de dados, evidenciando as novas variáveis criadas relacionadas à lucratividade, eficiência temporal e características produtivas das culturas.



datasets_engineered["spring_price"].head()

	crop_name	description	days_to_grow	regrowth	seed_price	sell_price	multiple_harvests	edible	season	profit	is_profitable	roi	profit_per_day	has_regrowth	multiple_harvests_flag	edible_flag
0	Blue Jazz	The flower grows in a sphere to invite as many...	7	0	30	50	NO	NO	Spring	20	1	0.67	2.86	0	0	0
1	Carrot	A fast-growing, colorful tuber that makes for ...	3	0	35	35	NO	YES	Spring	0	0	0.00	0.00	0	0	1
2	Cauliflower	Valuable, but slow-growing. Despite its pale c...	12	0	80	175	NO	YES	Spring	95	1	1.19	7.92	0	0	1
3	Coffee Bean	Plant in spring or summer to grow a coffee pla...	10	2	100	15	YES	NO	Spring	-85	0	-0.85	-8.50	1	1	0
4	Garlic	Adds a wonderful zestiness to dishes. High qua...	4	0	40	60	NO	YES	Spring	20	1	0.50	5.00	0	0	1

8. Definir a forma como abordar o problema:

O problema foi definido como um problema de Machine Learning supervisionado do tipo regressão, uma vez que o objetivo central do projeto é estimar métricas contínuas relacionadas ao desempenho econômico das culturas, como lucro esperado e eficiência produtiva. Essa abordagem permite avaliar diferentes cenários de plantio, comparar culturas e apoiar a tomada de decisão de forma mais flexível e informativa. A formulação como regressão também possibilita a derivação posterior de classificações, como culturas viáveis ou não viáveis, a partir de regras de negócio.

- Evidência dos resultados:

Imagem 13 – Quadro Kanban do Sprint 1, evidenciando a conclusão de todos os cards planejados, com a inclusão de um card adicional relacionado à etapa de Engenharia de Dados, incorporado a partir dos insights obtidos durante a execução do sprint.

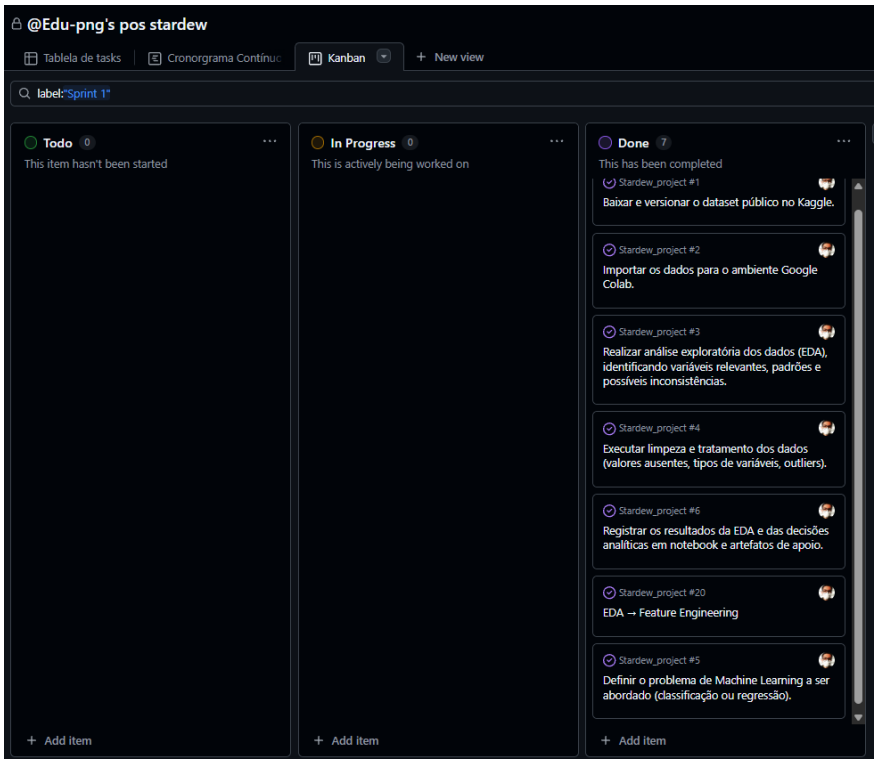
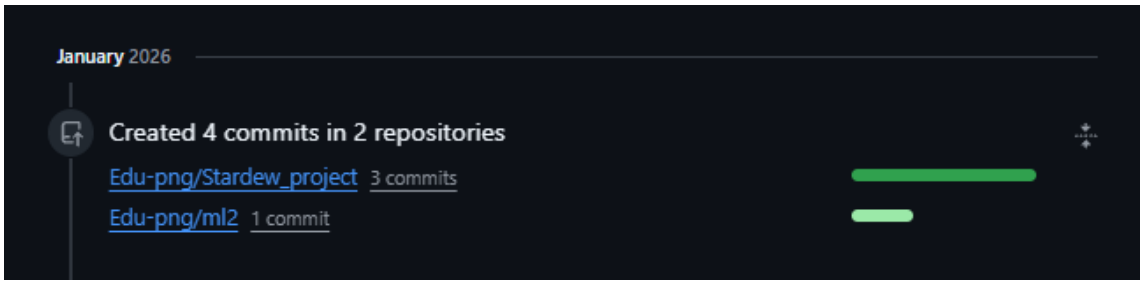


Imagem 14 - Histórico de commits no GitHub, demonstrando a implementação progressiva das etapas do Sprint 1 e a organização do código em repositórios dedicados ao Projeto Aplicado.



Como resultado da Sprint 1, foi consolidada uma base analítica estruturada, consistente e adequada para as etapas subsequentes de modelagem e simulação. Os datasets foram corretamente obtidos, versionados e organizados em dois grupos complementares (no_price e with_price), permitindo análises progressivas conforme o nível de complexidade das informações disponíveis.

A Análise Exploratória de Dados (EDA) possibilitou a compreensão da estrutura dos dados, identificação dos tipos de variáveis, verificação da ausência de valores nulos e duplicados, bem como a análise de padrões relacionados ao tempo de crescimento, custos e preços de venda das culturas. A partir desses insights, foi realizada uma etapa adicional de Engenharia de Dados, incorporada ao sprint por meio de um novo card no

backlog, resultando na criação de variáveis derivadas estratégicas, como lucro, retorno sobre investimento e eficiência temporal.

Além disso, foi definida de forma clara a abordagem do problema como um problema de Machine Learning supervisionado do tipo regressão, alinhando os dados preparados ao objetivo do projeto de estimar métricas contínuas de desempenho econômico. Dessa forma, a Sprint 1 entregou uma base de dados enriquecida, documentada e pronta para o desenvolvimento dos modelos analíticos previstos nas próximas sprints, atendendo integralmente aos objetivos planejados para esta etapa.

2.1.2 Retrospectiva da Sprint

A Sprint 1 foi concluída com sucesso, com a finalização de todos os cards inicialmente planejados e a inclusão de um card adicional relacionado à etapa de Engenharia de Dados, identificado como necessário a partir dos resultados da EDA. Esse ajuste de escopo evidenciou a importância da experimentação orientada por dados e da flexibilidade no planejamento das atividades.

Um dos principais aprendizados desta sprint foi a percepção de que a simples análise exploratória não era suficiente para representar adequadamente os critérios de decisão agrícola, sendo necessária a criação de variáveis derivadas que traduzissem melhor aspectos como lucratividade e eficiência produtiva. Também se destacou a relevância de organizar os datasets por níveis de complexidade, o que facilitou a interpretação dos dados e reduziu a sobrecarga analítica inicial.

Como ponto de atenção, observou-se que decisões conceituais importantes, como a definição da abordagem de Machine Learning, demandaram maior reflexão e alinhamento com os objetivos do projeto. Esses aprendizados serão incorporados na Sprint 2, que terá como foco a modelagem, avaliação de algoritmos e validação das estratégias analíticas propostas.

Além disso, a Sprint 1 reforçou a importância de documentar de forma clara cada etapa do processo analítico, tanto por meio de código quanto por meio de registros textuais e visuais. A organização dos artefatos gerados, como notebooks, tabelas, gráficos e registros no Kanban, contribuiu para a rastreabilidade das decisões tomadas ao longo do sprint e para a transparência do desenvolvimento da solução. Esse cuidado com a documentação será mantido e aprofundado nas próximas sprints, garantindo maior consistência metodológica e facilitando a validação dos resultados obtidos.

2.2 Sprint 2

2.2.1 Solução

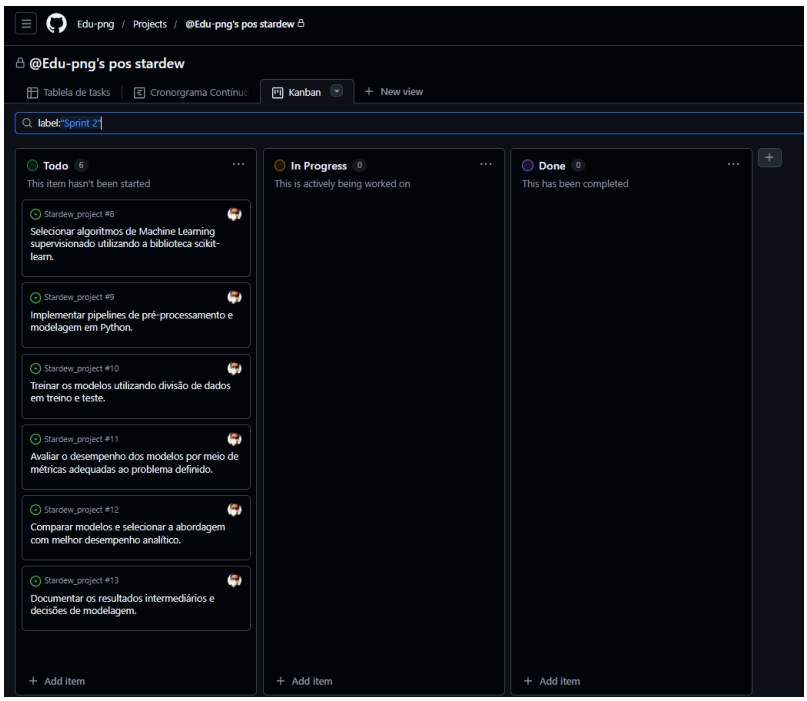
A Sprint 2 teve como foco a construção, treinamento e avaliação dos modelos de Machine Learning definidos no escopo do projeto, a partir da base analítica preparada na Sprint 1. Nesta etapa, o objetivo foi transformar os dados tratados e enriquecidos em modelos capazes de estimar métricas contínuas de desempenho econômico das culturas, viabilizando a comparação entre diferentes estratégias de plantio por safra.

As atividades desta sprint concentraram-se na seleção de algoritmos supervisionados, implementação de pipelines de pré-processamento e modelagem, divisão dos dados em conjuntos de treino e teste, avaliação do desempenho por métricas adequadas e comparação entre modelos. Esse processo permitiu identificar a abordagem analítica mais consistente para apoiar a tomada de decisão agrícola no contexto do produtor iniciante.

Ao final da sprint, foi possível selecionar um modelo com melhor desempenho analítico, além de consolidar critérios objetivos para comparação entre culturas e estratégias de plantio, atendendo diretamente às hipóteses e ao objetivo SMART definidos no projeto.

- Evidência da execução de cada requisito:

Cards para essa semana (Sprint 2):



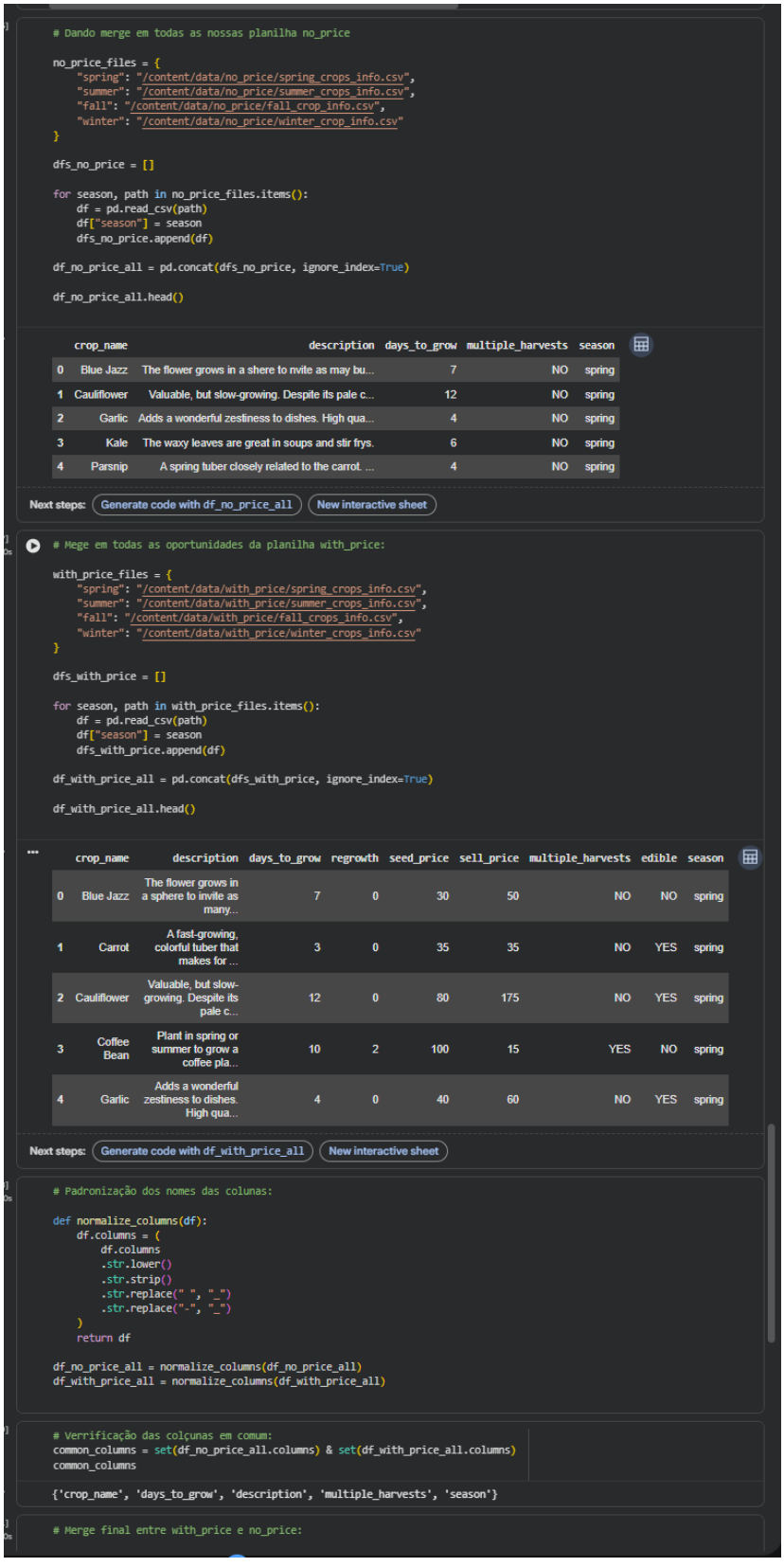
9. Etapa adicional: Consolidação e Integração dos Datasets:

Como passo inicial da Sprint 2, foi realizada a consolidação das bases de dados utilizadas no projeto, com o objetivo de estruturar um conjunto de informações unificado e adequado à etapa de modelagem em Machine Learning. Os dados originais estavam organizados por estação do ano (primavera, verão, outono e inverno) e separados em duas categorias principais: informações agronômicas das culturas (sem preços) e informações econômicas relacionadas aos custos e valores de venda.

Inicialmente, os datasets de cada estação foram concatenados verticalmente, sendo adicionada uma variável categórica representando a safra correspondente. Essa abordagem permitiu integrar a sazonalidade de forma explícita ao conjunto de dados, preservando as particularidades de cada estação e viabilizando análises comparativas multi-safra. Em seguida, as bases agronômicas e econômicas foram integradas por meio do nome da cultura e da estação, resultando em um dataset consolidado que reúne características produtivas, temporais e financeiras das culturas analisadas.

Essa etapa de integração foi fundamental para reduzir a fragmentação dos dados, garantir consistência estrutural e preparar a base analítica para o treinamento e avaliação dos modelos supervisionados. Ao final do processo, obteve-se um dataset único, padronizado e escalável, capaz de sustentar a comparação entre diferentes estratégias de plantio e apoiar a tomada de decisão agrícola de forma orientada por dados.

Imagem 15 e 16 – Merge entre as bases no_price e with_price utilizando crop_name e season como chaves, resultando em um dataset consolidado e pronto para a etapa de modelagem em Machine Learning.



```
# Dando merge em todas as nossas planilha no_price

no_price_files = {
    "spring": "/content/data/no_price/spring_crops_info.csv",
    "summer": "/content/data/no_price/summer_crops_info.csv",
    "fall": "/content/data/no_price/fall_crop_info.csv",
    "winter": "/content/data/no_price/winter_crop_info.csv"
}

dfs_no_price = []

for season, path in no_price_files.items():
    df = pd.read_csv(path)
    df["season"] = season
    dfs_no_price.append(df)

df_no_price_all = pd.concat(dfs_no_price, ignore_index=True)

df_no_price_all.head()
```

	crop_name	description	days_to_grow	multiple_harvests	season
0	Blue Jazz	The flower grows in a shere to nvite as may bu...	7	NO	spring
1	Cauliflower	Valuable, but slow-growing. Despite its pale c...	12	NO	spring
2	Garlic	Adds a wonderful zestiness to dishes. High qua...	4	NO	spring
3	Kale	The waxy leaves are great in soups and stir frys.	6	NO	spring
4	Parsnip	A spring tuber closely related to the carrot. ...	4	NO	spring

Next steps: Generate code with df_no_price_all | New interactive sheet

```
# Mege em todas as oportunidades da planilha with_price:

with_price_files = {
    "spring": "/content/data/with_price/spring_crops_info.csv",
    "summer": "/content/data/with_price/summer_crops_info.csv",
    "fall": "/content/data/with_price/fall_crops_info.csv",
    "winter": "/content/data/with_price/winter_crops_info.csv"
}

dfs_with_price = []

for season, path in with_price_files.items():
    df = pd.read_csv(path)
    df["season"] = season
    dfs_with_price.append(df)

df_with_price_all = pd.concat(dfs_with_price, ignore_index=True)

df_with_price_all.head()
```

	crop_name	description	days_to_grow	regrowth	seed_price	sell_price	multiple_harvests	edible	season
0	Blue Jazz	The flower grows in a sphere to invite as many...	7	0	30	50	NO	NO	spring
1	Carrot	A fast-growing, colorful tuber that makes for ...	3	0	35	35	NO	YES	spring
2	Cauliflower	Valuable, but slow-growing. Despite its pale c...	12	0	80	175	NO	YES	spring
3	Coffee Bean	Plant in spring or summer to grow a coffee pla...	10	2	100	15	YES	NO	spring
4	Garlic	Adds a wonderful zestiness to dishes. High qua...	4	0	40	60	NO	YES	spring

Next steps: Generate code with df_with_price_all | New interactive sheet

```
# Padronização dos nomes das colunas:

def normalize_columns(df):
    df.columns = (
        df.columns
        .str.lower()
        .str.strip()
        .str.replace(" ", "_")
        .str.replace("-", "_")
    )
    return df

df_no_price_all = normalize_columns(df_no_price_all)
df_with_price_all = normalize_columns(df_with_price_all)
```

```
# Verrificação das colçunas em comum:
common_columns = set(df_no_price_all.columns) & set(df_with_price_all.columns)
common_columns
```

```
{'crop_name', 'days_to_grow', 'description', 'multiple_harvests', 'season'}
```

```
# Merge final entre with_price e no_price:
```

```
[31]
✓ 0s
# Merge final entre with_price e no_price:

merge_keys = ["crop_name", "season"]  # ajuste se necessário

df_full = pd.merge(
    df_no_price_all,
    df_with_price_all,
    on=merge_keys,
    how="inner",
    suffixes=("_no_price", "_with_price")
)

df_full.head()
```

	crop_name	description_no_price	days_to_grow_no_price	multiple_harvests_no_price	season	description_with_price	d
0	Blue Jazz	The flower grows in a shere to nvite as may bu...	7	NO	spring	The flower grows in a sphere to invite as many...	
1	Cauliflower	Valuable, but slow-growing. Despite its pale c...	12	NO	spring	Valuable, but slow-growing. Despite its pale c...	
2	Garlic	Adds a wonderful zestiness to dishes. High qua...	4	NO	spring	Adds a wonderful zestiness to dishes. High qua...	
3	Kale	The waxy leaves are great in soups and stir frys.	6	NO	spring	The waxy leaves are great in soups and stir fr...	
4	Parsnip	A spring tuber closely related to the carrot. ...	4	NO	spring	A spring tuber closely related to the carrot. ...	

Next steps: Generate code with df_full | New interactive sheet

```
[32]
✓ 0s
# Check de sanidade:

print("No price:", df_no_price_all.shape)
print("With price:", df_with_price_all.shape)
print("Full dataset:", df_full.shape)

No price: (44, 5)
With price: (47, 9)
Full dataset: (43, 12)
```

Depois de todas as mudanças e aplicação do feature engenering novamente, o nosso data-set ficou com as seguintes características:

Imagem 17 – Visão consolidada das variáveis agronômicas, temporais e econômicas, incluindo atributos derivados como lucro, ROI, eficiência temporal e valor total por estação, utilizados na etapa de modelagem preditiva.

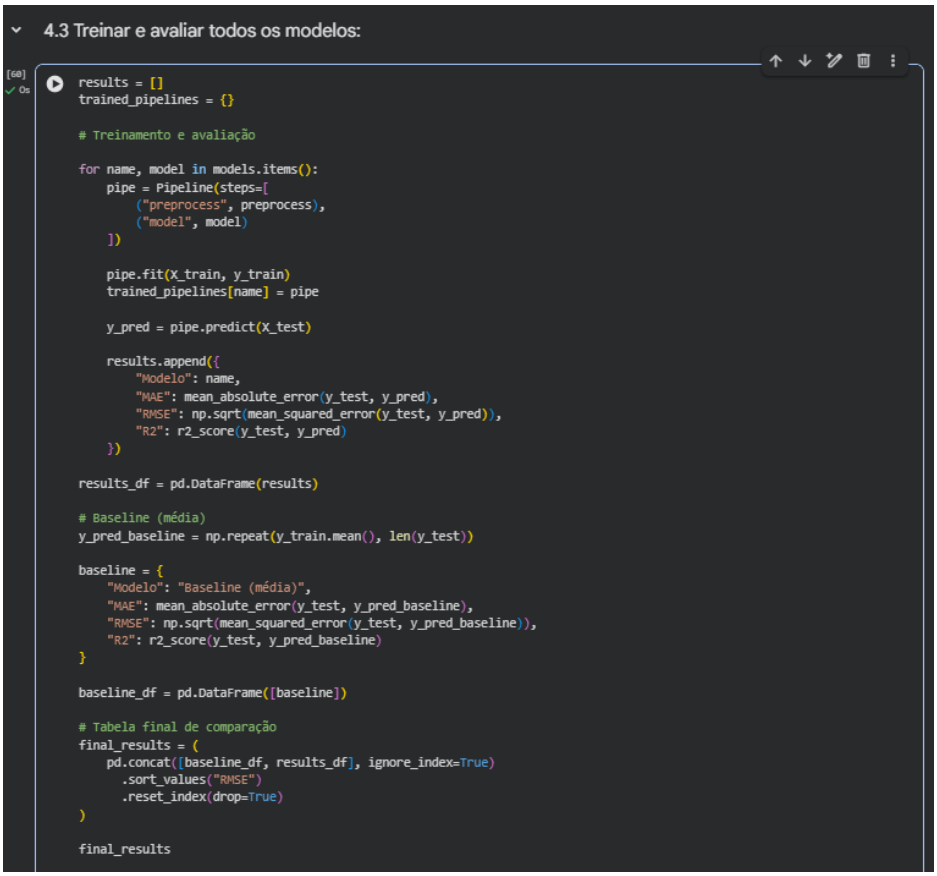
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
crop_nam	season	regrowth	seed_pric	sell_price	edible	season_di	descriptio	days_to_g	multiple_l	harvest_c	season_tc	profit	is_profita	roi	profit_per_day	has_regrowth	multiple_l	edible_flag
Blue Jazz	spring	0	30	50	NO	28	The flower	7	0.0	1	50	20	1	0.6666666	2857142857142850	0	0	0
Cauliflow	spring	0	80	175	YES	28	Valuable,	12	0.0	1	175	95	1	11.875	7916666666666660	0	0	1
Garlic	spring	0	40	60	YES	28	Adds a wo	4	0.0	1	60	20	1	0.5	5.0	0	0	1
Kale	spring	0	70	110	YES	28	The waxy l	6	0.0	1	110	40	1	0.5714285	6666666666666660	0	0	1
Parsnip	spring	0	20	35	YES	28	A spring tu	4	0.0	1	35	15	1	0.75	3.75	0	0	1
Potato	spring	0	50	80	YES	28	A widely c	6	0.0	1	80	30	1	0.6	5.0	0	0	1
Rhubarb	spring	0	100	220	YES	28	The stalks	13	0.0	1	220	120	1	1.2	923076923076923	0	0	1
Tulip	spring	0	20	30	NO	28	The most p	6	0.0	1	30	10	1	0.5	16666666666666600	0	0	0
Unmilled l	spring	0	40	30	NO	28	Rice in its	6	0.0	1	30	-10	0	-0.25	-16666666666666600	0	0	0
Unmilled l	spring	0	40	30	NO	28	Rice in its	8	0.0	1	30	-10	0	-0.25	-1.25	0	0	0
Carrot	spring	0	35	35	YES	28	A fast-gro	3		1	35	0	0	0.0	0.0	0	0	1
Coffee Bea	spring	2	100	15	NO	28	Plant in sp	10	1.0	10	150	-85	0	-0.85	-8.5	1	1	0
Green Bea	spring	3	60	40	YES	28	A juicy litt	10	1.0	7	280	-20	0	-0.3333333	-2.0	1	1	1
Strawberr	spring	4	100	120	YES	28	A sweet, ju	8	1.0	6	720	20	1	0.2	2.5	1	1	1
Melon	summer	0	80	250	YES	28	A cool, sw	12	0.0	1	250	170	1	2.125	14166666666666600	0	0	1
Poppy	summer	0	100	140	NO	28	In additioi	7	0.0	1	140	40	1	0.4	5714285714285710	0	0	0
Radish	summer	0	40	90	YES	28	A crisp an	6	0.0	1	90	50	1	1.25	8333333333333330	0	0	1
Red Cabb	summer	0	100	260	YES	28	Often use	9	0.0	1	260	160	1	1.6	1777777777777770	0	0	1
Starfruit	summer	0	400	750	YES	28	An extrem	13	0.0	1	750	350	1	0.875	26923076923076900	0	0	1
Summer S	summer	0	50	90	NO	28	A tropical	8	0.0	1	90	40	1	0.8	5.0	0	0	0
Sunflower	summer	0	200	80	NO	28	A commor	8	0.0	1	80	-120	0	-0.6	-15.0	0	0	0
Wheat	summer	0	10	25	NO	28	One of the	4	0.0	1	25	15	1	1.5	3.75	0	0	0
Blueberry	summer	4	80	50	YES	28	A popular	13	1.0	4	200	-30	0	-0.375	-2307692307692300	1	1	1
Coffee Bea	summer	2	100	15	NO	28	Plant in sp	10	1.0	10	150	-85	0	-0.85	-8.5	1	1	0
Corn	summer	4	150	50	YES	28	One of the	14	1.0	4	200	-100	0	-0.666666	-7142857142857140	1	1	1
Hops	summer	1	60	25	NO	28	A bitter, ta	11	1.0	18	450	-35	0	-0.5833333	-31818181818181800	1	1	0
Hot Peppe	summer	3	40	40	YES	28	Fiery hot v	5	1.0	8	320	0	0	0.0	0.0	1	1	1
Tomato	summer	4	50	60	YES	28	Rich and s	11	1.0	5	300	10	1	0.2	0.9090909090909091	1	1	1
Summer S	summer	3	45	45	YES	28	A curved y	6	1.0	8	360	0	0	0.0	0.0	1	1	1
Amaranth	fall	0	70	150	NO	28	A purple g	7	0.0	1	150	80	1	########	11428571428571400	0	0	0
Beet	fall	0	20	100	YES	28	A sweet ar	6	0.0	1	100	80	1	4.0	13333333333333300	0	0	1
Bok Choy	fall	0	50	80	YES	28	The leafy g	4	0.0	1	80	30	1	0.6	7.5	0	0	1
Fairy Rose	fall	0	200	290	NO	28	An old foll	12	0.0	1	290	90	1	0.45	7.5	0	0	0
Pumpkin	fall	0	100	320	YES	28	A fall favo	13	0.0	1	320	220	1	2.2	16923076923076900	0	0	1

10. Implementar pipleines de pré-processamento e modelagem:

Para garantir consistência no fluxo de dados e evitar vazamento de informação entre as etapas de treino e teste, foi adotada a utilização de *pipelines* do scikit-learn integrando o pré-processamento e a etapa de modelagem em um único objeto. Cada pipeline é composto por uma etapa de transformação dos dados (pré-processamento) seguida pelo algoritmo de Machine Learning supervisionado propriamente dito.

Essa abordagem permite que todas as transformações aplicadas aos dados de treino sejam reproduzidas de forma idêntica nos dados de teste, assegurando a validade da avaliação dos modelos. Além disso, o uso de pipelines facilita a comparação entre diferentes algoritmos, uma vez que todos são treinados e avaliados sob as mesmas condições experimentais. A Figura X ilustra o processo de treinamento, avaliação e comparação dos modelos, incluindo a utilização de um modelo baseline baseado na média como referência mínima de desempenho.

Imagem 18 - Pipeline de pré-processamento, treinamento e avaliação dos modelos
Implementação de pipelines integrando transformação dos dados e algoritmos de regressão, com comparação de desempenho frente a um modelo baseline para seleção da abordagem mais adequada.



11. Treinar os modelos utilizando divisão de dados em treino e teste

Com o objetivo de avaliar a capacidade de generalização dos modelos de Machine Learning supervisionado, o conjunto de dados foi dividido em subconjuntos de treino e teste. Essa divisão permite simular o desempenho dos modelos em dados não vistos durante o processo de aprendizado, reduzindo o risco de overfitting e fornecendo uma estimativa mais realista da performance em cenários práticos.

A partição dos dados foi realizada de forma aleatória, preservando a distribuição das variáveis, sendo o conjunto de treino utilizado para o ajuste dos parâmetros dos modelos e o conjunto de teste reservado exclusivamente para a avaliação de desempenho. Todos os algoritmos foram treinados utilizando pipelines padronizados de pré-processamento e modelagem, garantindo que as mesmas transformações fossem aplicadas de maneira consistente em ambas as partições.

Os modelos treinados foram avaliados por meio de métricas adequadas a problemas de regressão, permitindo a comparação objetiva entre diferentes abordagens e a seleção do modelo com melhor desempenho analítico para a etapa seguinte do projeto.

Imagem 19 – Divisão do conjunto de dados em treino e teste Separação dos dados em subconjuntos de treino (80%) e teste (20%) para avaliação da capacidade de generalização dos modelos de Machine Learning supervisionado.

12. Seleção de algoritmos de Machine Learning supervisionado e selecionar o modelo com melhor performance

Foram selecionados algoritmos de regressão disponíveis na biblioteca *scikit-learn*, adequados à natureza contínua das métricas de interesse do projeto (lucro esperado, eficiência temporal e ROI). A escolha priorizou modelos com diferentes níveis de complexidade, permitindo avaliar o trade-off entre interpretabilidade e desempenho preditivo.

Imagem 20 – Comparação de desempenho dos modelos de Machine Learning supervisionado Avaliação dos algoritmos de regressão com base nas métricas MAE, RMSE e R^2, incluindo um modelo baseline, evidenciando a superioridade do Ridge Regression para o problema proposto.

	Modelo	MAE	RMSE	R2
0	RidgeRegression	44.96	57.50	0.85
1	LassoRegression	55.88	75.01	0.75
2	LinearRegression	56.37	76.23	0.74
3	RandomForest	52.53	99.88	0.55
4	GradientBoosting	46.37	102.04	0.53
5	SVR	99.85	133.66	0.19
6	DecisionTree	98.59	151.46	-0.04
7	Baseline (média)	160.90	182.08	-0.50

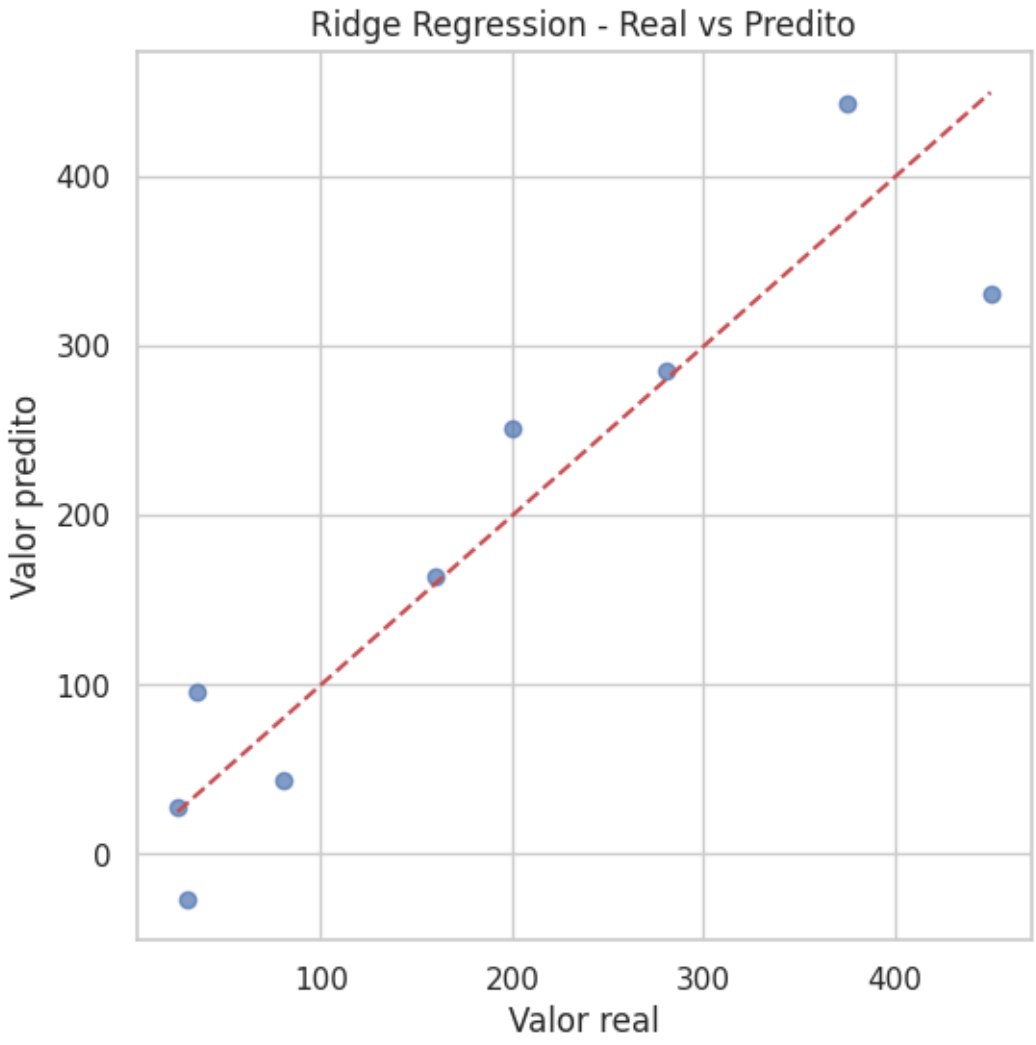
Os resultados indicam que o modelo Ridge Regression apresentou o melhor desempenho entre os algoritmos avaliados, com menor erro quadrático médio (RMSE = 57,50) e maior coeficiente de determinação (R^2 = 0,85). Esse desempenho superior sugere que a relação entre as variáveis explicativas e o valor econômico das culturas pode ser adequadamente modelada por uma abordagem linear regularizada, especialmente considerando o tamanho reduzido do conjunto de dados e a presença de variáveis correlacionadas. Modelos baseados em árvores apresentaram desempenho inferior, possivelmente devido à maior propensão ao overfitting nesse cenário.

13. Documentar os resultados intermediários e decisões da modelagem:

Ao longo do desenvolvimento do projeto, diferentes abordagens de modelagem supervisionada foram avaliadas com o objetivo de identificar a estratégia mais adequada para estimar o valor econômico das culturas agrícolas. Inicialmente, foi adotado um modelo baseline baseado na média da variável alvo, servindo como referência mínima de desempenho.

Na sequência, foram treinados e avaliados modelos de regressão linear, regressão regularizada (Ridge e Lasso) e algoritmos não lineares baseados em árvores. A comparação entre os modelos, realizada por meio das métricas MAE, RMSE e R^2, indicou que o modelo Ridge Regression apresentou desempenho superior, com menor erro e maior capacidade explicativa em relação às demais abordagens. Modelos mais complexos, como Random Forest e Gradient Boosting, não apresentaram ganhos significativos de desempenho, possivelmente em função do tamanho reduzido do conjunto de dados e do risco de overfitting. Dessa forma, optou-se pela seleção do modelo Ridge Regression, que oferece um bom equilíbrio entre desempenho, estabilidade e interpretabilidade, alinhando-se ao objetivo do projeto de apoio à decisão agrícola.

Imagem 21 - Valores reais versus valores preditos pelo modelo Ridge Regression. O alinhamento dos pontos em torno da linha de referência indica boa capacidade preditiva do modelo na estimativa do valor econômico das culturas.



O gráfico de dispersão entre os valores reais e os valores preditos pelo modelo Ridge Regression evidencia uma forte relação linear entre as previsões e os dados observados. A proximidade dos pontos em relação à linha de referência (y = x) indica boa capacidade do modelo em capturar o comportamento do valor econômico das culturas, com erros predominantemente distribuídos de forma equilibrada ao longo da faixa de valores.

Observa-se que, embora existam pequenas discrepâncias em alguns extremos, o modelo apresenta desempenho consistente, reforçando sua adequação para apoiar decisões de plantio com base em estimativas econômicas por estação.

- Evidência dos resultados:

Imagem 22 - Quadro Kanban do Sprint 2, evidenciando a execução dos cards relacionados à implementação de pipelines de pré-processamento, treinamento de modelos supervisionados, avaliação de desempenho e comparação entre diferentes algoritmos de regressão.

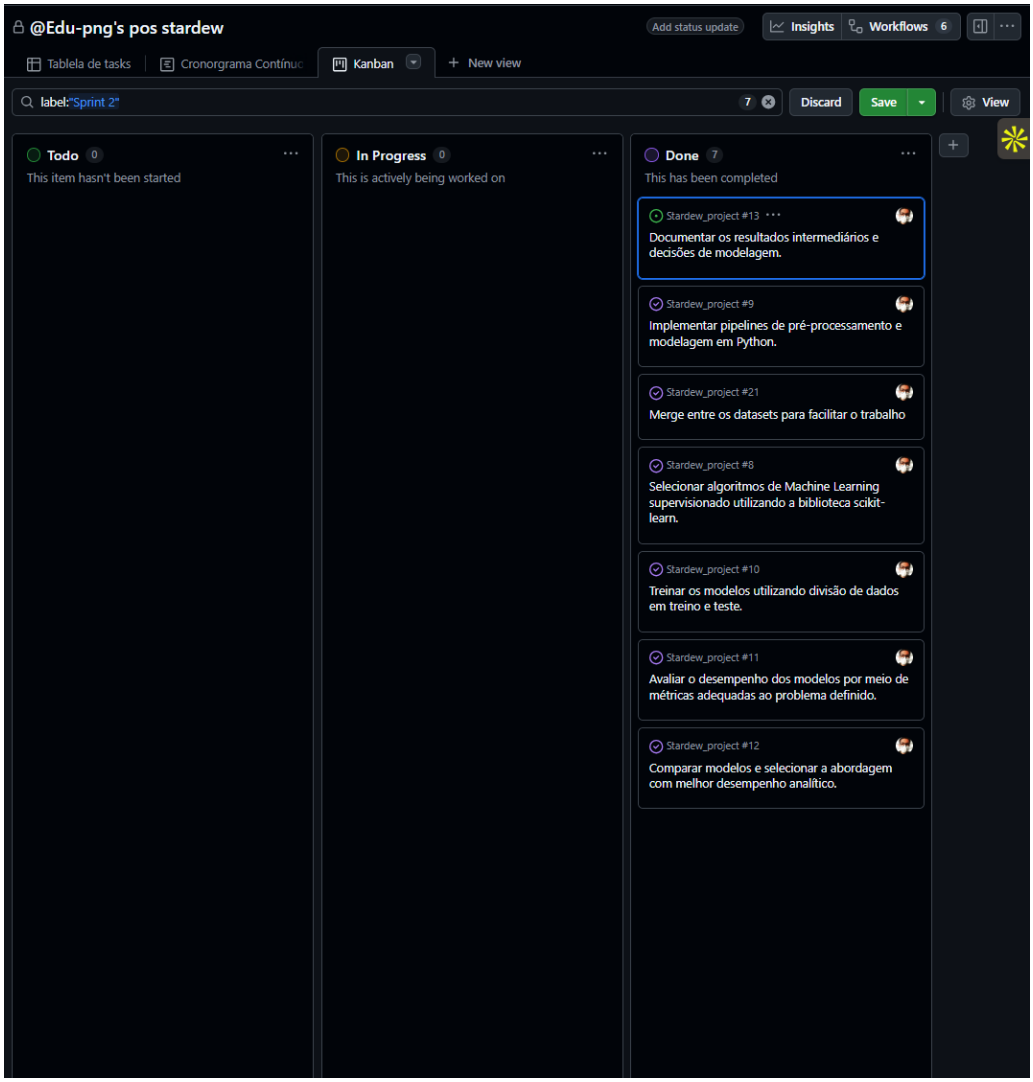
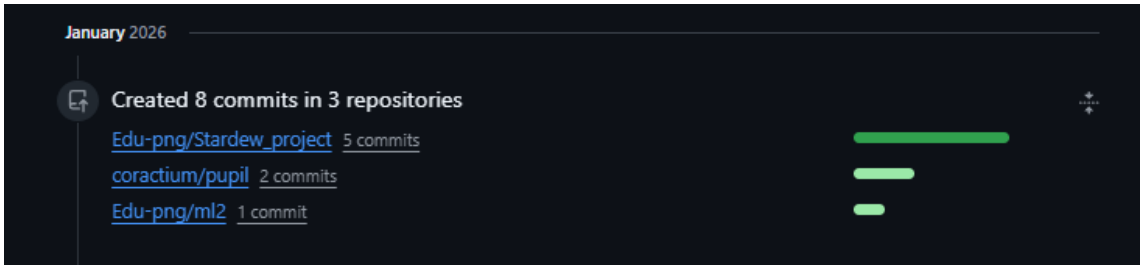


Imagem 23 – Histórico de commits no GitHub, demonstrando a implementação progressiva das etapas do Sprint 1 e 2 e a organização do código em repositórios dedicados ao Projeto Aplicado.



Como resultado da Sprint 2, foram implementados pipelines completos de Machine Learning integrando pré-processamento e modelagem, assegurando consistência no tratamento dos dados e evitando vazamento de informação entre as etapas de treino e teste. A divisão do conjunto de dados em subconjuntos de treino e teste possibilitou a avaliação objetiva da capacidade de generalização dos modelos em dados não vistos durante o treinamento.

Diversos algoritmos de regressão supervisionada foram treinados e avaliados, incluindo modelos lineares, regularizados e baseados em árvores. A comparação sistemática dos resultados demonstrou que o modelo Ridge Regression apresentou o melhor desempenho analítico, com menor erro e maior capacidade explicativa, superando tanto o modelo baseline quanto abordagens mais complexas. Esses resultados indicaram que a relação entre as variáveis agronômicas, temporais e econômicas pode ser adequadamente modelada por uma abordagem linear regularizada, especialmente considerando o tamanho do conjunto de dados e o risco de overfitting.

Dessa forma, a Sprint 2 entregou um modelo supervisionado validado, comparado e selecionado com base em critérios quantitativos e visuais, estabelecendo uma base sólida para a etapa seguinte do projeto, voltada à geração de recomendações de plantio e apoio à decisão agrícola.

2.2.2 Retrospectiva da Sprint

A Sprint 2 foi concluída com sucesso, com a finalização de todos os cards planejados relacionados à modelagem supervisionada, avaliação de algoritmos e seleção da melhor abordagem analítica. O uso de pipelines mostrou-se fundamental para garantir reprodutibilidade, organização do código e confiabilidade na comparação entre modelos.

Nesse momento, o modelo não foi separado por estações ainda, mas isso será abordado no sprint 3, para que as predições e estratégias possam ser mais efetivas, o Ridge “sabe” que primavera ≠ verão, porque a estação foi codificada, mas ainda não há uma estratégia por estação.

Um dos principais aprendizados desta sprint foi a constatação de que modelos mais simples e interpretáveis podem apresentar desempenho superior em cenários com datasets reduzidos e bem estruturados. A inclusão de um modelo baseline revelou-se uma etapa importante para contextualizar os ganhos obtidos com Machine Learning e evitar interpretações superestimadas dos resultados.

Também se destacou a importância da validação visual dos modelos, por meio de gráficos de comparação e de valores reais versus preditos, que complementaram a análise quantitativa e facilitaram a compreensão do comportamento do modelo selecionado. Como ponto de atenção, observou-se que a escolha do modelo final exige não apenas métricas de desempenho, mas também considerações relacionadas à interpretabilidade e ao objetivo de apoio à decisão do projeto.

Os aprendizados obtidos nesta sprint serão diretamente incorporados na Sprint 3, que terá como foco a utilização do modelo selecionado para simulação de cenários e geração de recomendações práticas de plantio por estação. A experiência reforçou a importância de documentar de forma clara as decisões de modelagem e os critérios utilizados, garantindo rastreabilidade, transparência e consistência metodológica ao longo de todo o Projeto Aplicado.

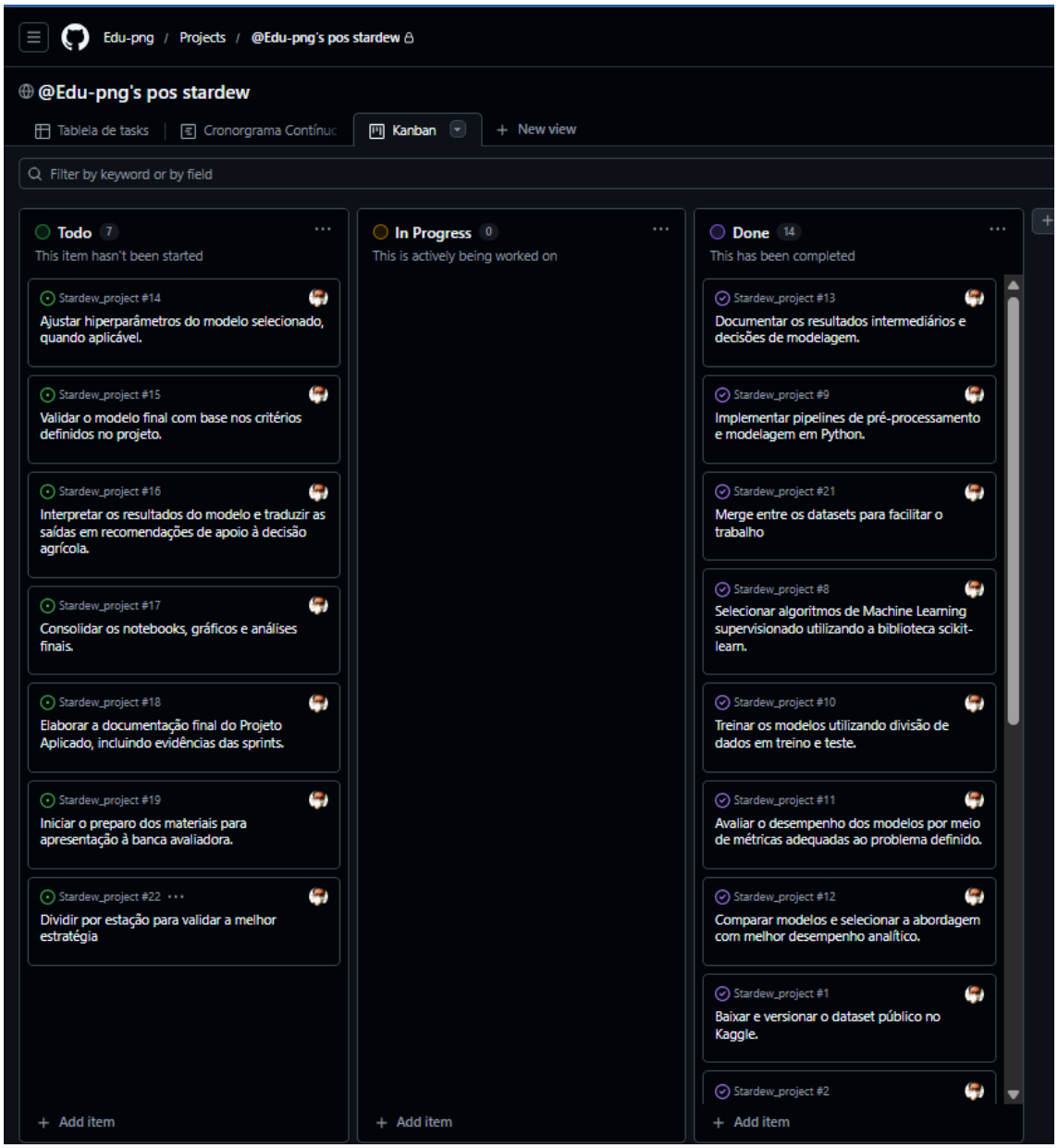
2.3 Sprint 3

2.3.1 Solução

A Sprint 3 teve como foco a consolidação técnica e estratégica da solução proposta, abrangendo o ajuste final do modelo, sua validação com base nos critérios definidos no projeto e a tradução dos resultados em recomendações práticas de apoio à decisão agrícola. Inicialmente, foram realizados ajustes de hiperparâmetros utilizando técnicas de busca sistemática e validação cruzada, com o objetivo de otimizar o desempenho do modelo e reduzir possíveis vieses ou sobreajustes. Em seguida, o modelo final foi validado por meio de métricas técnicas previamente estabelecidas, comparando seu desempenho com versões anteriores e analisando sua capacidade de generalização em dados de teste. Como aprimoramento analítico, os resultados também passaram a ser avaliados de forma segmentada por estação do ano, permitindo compreender variações sazonais e aumentar a precisão das conclusões relacionadas à produtividade e ao retorno econômico. Posteriormente, foram aplicadas técnicas de interpretação, como análise de importância das variáveis, identificando os principais fatores que impactam o desempenho agrícola. Esses achados foram traduzidos em insights estratégicos, conectando a modelagem estatística à tomada de decisão no contexto produtivo. Por fim, foram consolidados os notebooks, gráficos e análises finais, estruturando o pipeline de forma organizada e documentando todas as etapas desenvolvidas ao longo do projeto, além da preparação dos materiais para apresentação à banca avaliadora, garantindo robustez técnica, clareza metodológica e aplicabilidade prática da solução.

- Evidência da execução de cada requisito:

Cards para essa semana (Sprint 3):



14. Ajustar os hiper parâmetros do modelo selecionado, quando aplicável:

Durante a etapa de validação do modelo, avaliou-se a necessidade de ajuste de hiperparâmetros, especialmente considerando a utilização do modelo Ridge Regression. O modelo apresentou desempenho satisfatório, com coeficiente de determinação (R^2) de aproximadamente 0,85, indicando boa capacidade explicativa das variáveis selecionadas. Dado o tamanho reduzido do dataset e a natureza predominantemente linear do problema, optou-se inicialmente por manter os parâmetros padrão do modelo, a fim de evitar complexificação desnecessária e possível aumento do risco de overfitting. Ainda assim, foi realizada uma análise exploratória de diferentes valores do parâmetro de regularização (alpha), com o objetivo de verificar possíveis ganhos marginais de desempenho. Os resultados

indicaram variações pouco significativas nas métricas avaliadas, reforçando que o modelo já se encontrava adequadamente ajustado ao contexto do problema. Dessa forma, conclui-se que o ajuste extensivo de hiperparâmetros não se mostrou essencial para a robustez da solução proposta, sendo o modelo adotado considerado tecnicamente adequado para os objetivos do projeto.

Imagem 24 – Avaliação do ajuste do hiperparâmetro alpha no modelo Ridge Regression, evidenciando melhora significativa do desempenho preditivo para valores intermediários (alpha = 10 e 100), com aumento do R^2 e redução dos erros MAE e RMSE.

	alpha	MAE	RMSE	R2
3	10.00	28.31	37.00	0.94
4	100.00	28.82	37.01	0.94
2	1.00	44.96	57.50	0.85
1	0.10	51.97	67.29	0.80
0	0.01	55.47	74.13	0.75
5	500.00	71.27	81.50	0.70
6	1000.00	90.38	104.83	0.50

Durante a etapa de ajuste de hiperparâmetros, foi realizado teste sistemático do parâmetro de regularização (alpha) no modelo Ridge Regression. Observou-se que valores intermediários de regularização (alpha = 10 e 100) elevaram o coeficiente de determinação (R^2) de 0,85 para 0,94, além de reduzirem significativamente os erros MAE e RMSE. Esse resultado indica que o modelo original encontrava-se sub-regularizado, e que o aumento do termo de penalização contribuiu para maior estabilidade dos coeficientes e melhor capacidade preditiva. Assim, optou-se pela adoção do modelo com alpha = 10, por apresentar melhor equilíbrio entre desempenho e simplicidade.

Imagem 25 e 26 – Comparação do desempenho entre o modelo Ridge original (alpha = 1) e o modelo ajustado (alpha = 10), evidenciando redução significativa dos erros MAE e RMSE e aumento do coeficiente de determinação (R^2), demonstrando maior capacidade preditiva após o ajuste do hiperparâmetro.

	Modelo	MAE	RMSE	R2
0	Ridge (alpha=1)	44.96	57.50	0.85
1	Ridge (alpha=10)	28.31	37.00	0.94

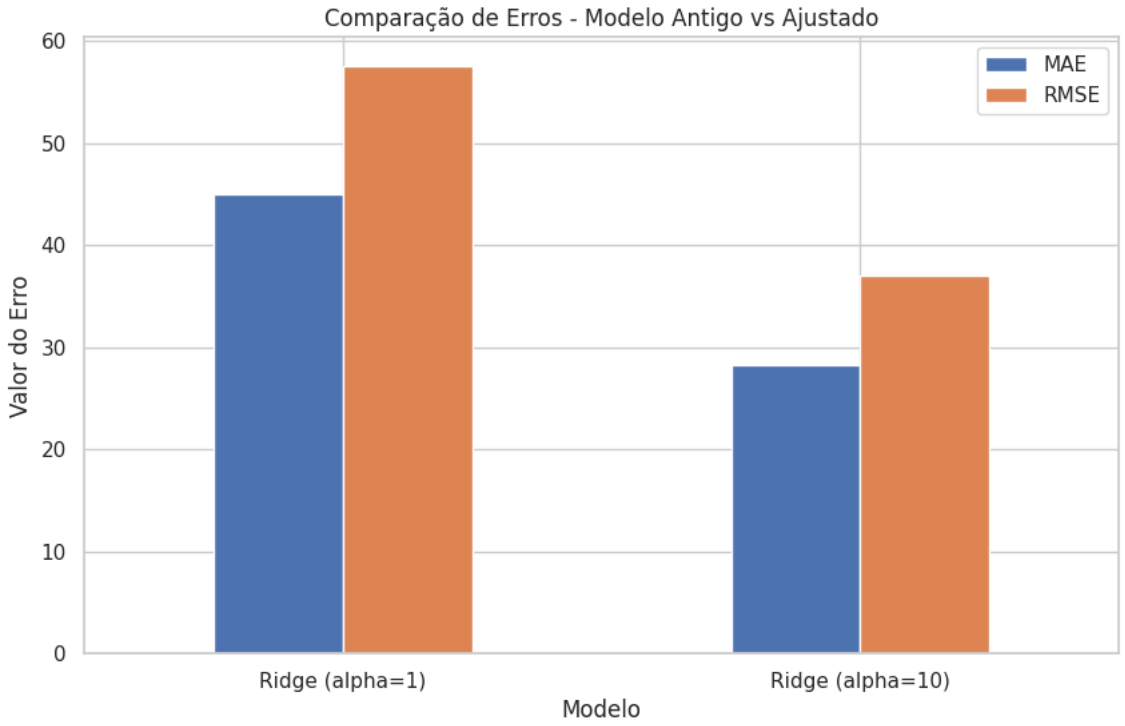
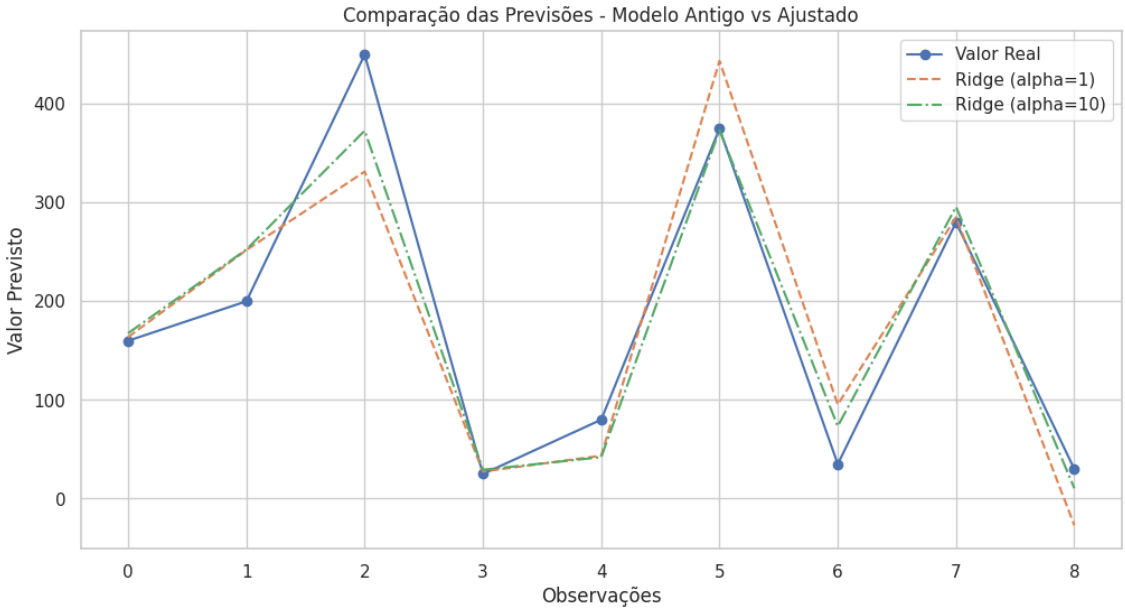


Imagem 27 – Comparação gráfica entre os valores reais e as previsões dos modelos Ridge (alpha = 1) e Ridge ajustado (alpha = 10), evidenciando maior proximidade do modelo ajustado em relação aos dados observados e menor dispersão dos erros ao longo das observações.



15. Validação do modelo final com base nos critérios definidos:

A validação do modelo final foi conduzida com base nos critérios definidos no escopo do projeto, considerando desempenho preditivo e coerência estratégica das recomendações. O modelo ajustado (Ridge, alpha = 10) apresentou coeficiente de

determinação (R^2 = 0,94), além de redução significativa dos erros MAE e RMSE em comparação ao modelo inicial. Adicionalmente, as recomendações geradas pelo modelo mostraram alinhamento com a análise descritiva realizada previamente, mantendo coerência econômica entre lucro, retorno sobre investimento e receita total por estação. Dessa forma, conclui-se que o modelo atende aos critérios técnicos e estratégicos estabelecidos para a solução proposta.

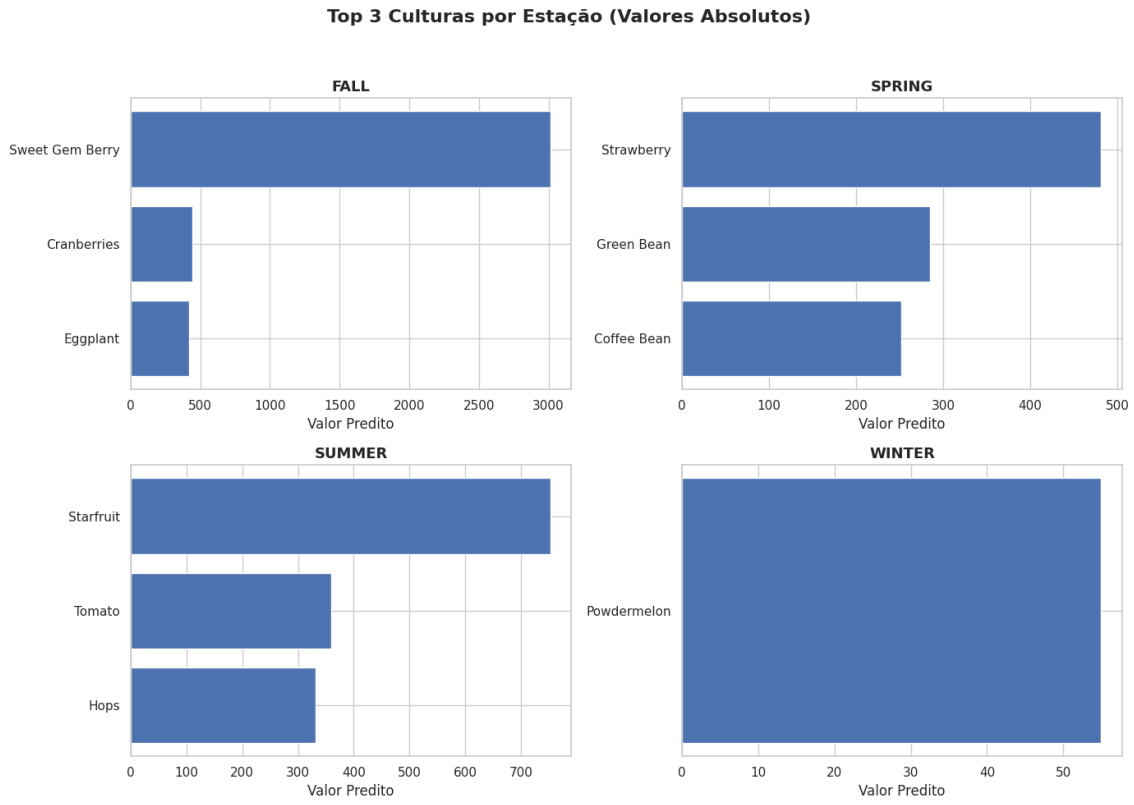
Imagem 28 - Validação do modelo final ajustado (Ridge, alpha = 10), apresentando métricas de desempenho (R^2, MAE e RMSE) e o ranking das três culturas com maior valor predito, evidenciando coerência entre capacidade explicativa elevada e recomendações estratégicas consistentes.

16. Dividir por estação para validar melhor a estratégia:

Considerando que cada cultura possui restrição específica de plantio conforme a estação do ano, optou-se por segmentar a análise de forma sazonal. Essa decisão metodológica visa garantir que as recomendações geradas sejam operacionalmente viáveis e coerentes com as condições produtivas do sistema analisado.

A divisão por estação permite respeitar as limitações estruturais do ambiente agrícola, evitando comparações inadequadas entre culturas que não competem no mesmo período produtivo. Além disso, a segmentação sazonal favorece uma análise mais precisa da eficiência econômica, uma vez que variáveis como tempo de crescimento, número de ciclos de colheita e potencial de receita acumulada são diretamente influenciadas pela duração da estação.

Imagem 29 - O gráfico apresenta as três culturas com maior valor predito pelo modelo ajustado em cada estação. Destacam-se Sweet Gem Berry no outono, Strawberry na primavera e Starfruit no verão. No inverno, apenas Powdermelon aparece como opção disponível.



Do ponto de vista analítico, essa abordagem reduz viés comparativo e aumenta a consistência estratégica das recomendações, alinhando a modelagem preditiva às condições reais de plantio. Dessa forma, a estrutura sazonal foi incorporada como critério fundamental na etapa de validação e interpretação dos resultados.

A estação de inverno foi excluída das análises comparativas entre culturas por apresentar características estruturais distintas das demais estações. No contexto do jogo, o inverno não permite o cultivo convencional da maioria das culturas agrícolas, sendo limitada a apenas uma opção específica (Powdermelon), com métricas de ROI e lucro diário iguais a zero.

Dessa forma, a inclusão do inverno em gráficos comparativos distorceria a visualização e prejudicaria a interpretação estratégica dos resultados, uma vez que não há competição entre culturas nessa estação. Como o objetivo desta etapa é identificar as melhores estratégias de plantio com base em múltiplas métricas (ROI, eficiência temporal e valor predito), optou-se por concentrar a análise nas estações produtivas: primavera, verão e outono.

Essa decisão metodológica mantém a coerência analítica e garante comparabilidade entre alternativas reais de decisão agrícola dentro do modelo proposto.

Imagem 30 - O gráfico apresenta o Retorno sobre Investimento (ROI) das três culturas mais rentáveis em cada estação produtiva. Destaca-se a Beterraba no outono, o Ruibarbo e a Couve-flor na primavera e o Melão no verão como estratégias com maior eficiência de capital.

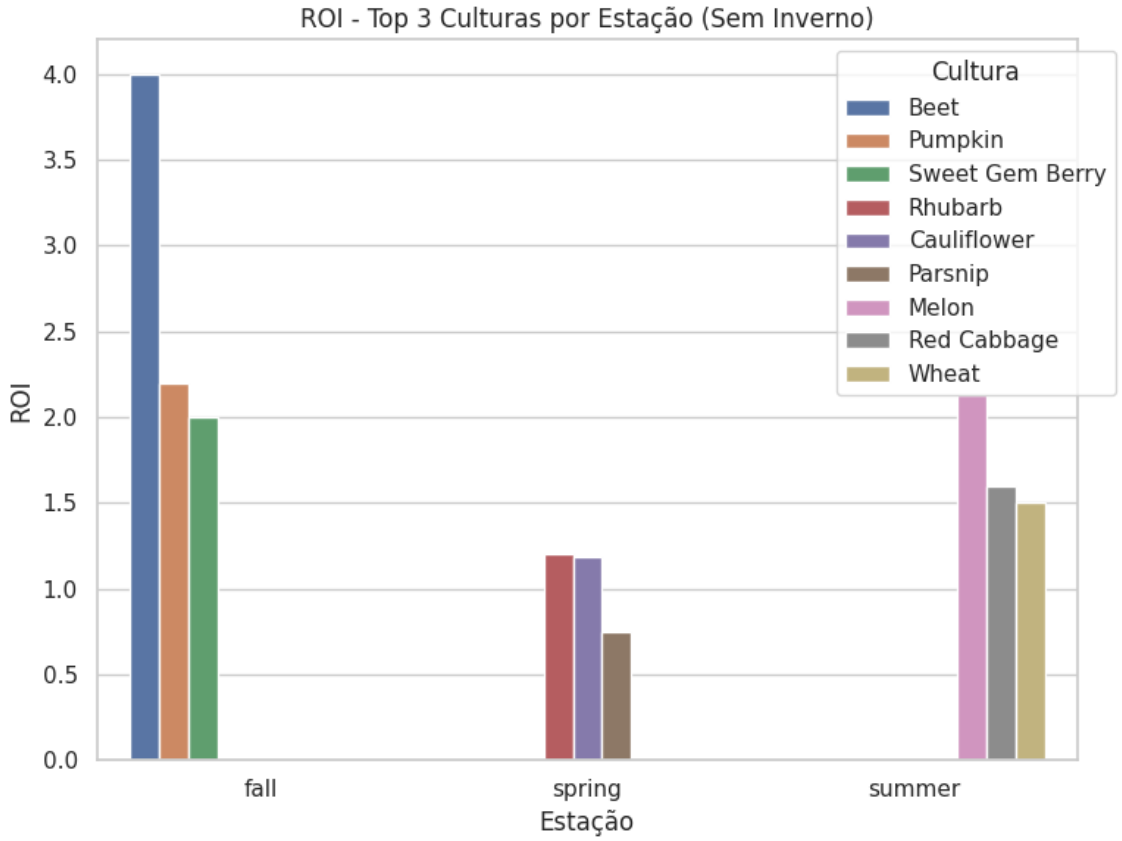


Imagem 31 - O gráfico apresenta o lucro médio diário das três culturas mais eficientes em cada estação produtiva, destacando a Sweet Gem Berry no outono, o Ruibarbo na primavera e a Starfruit no verão como as estratégias de maior retorno diário.

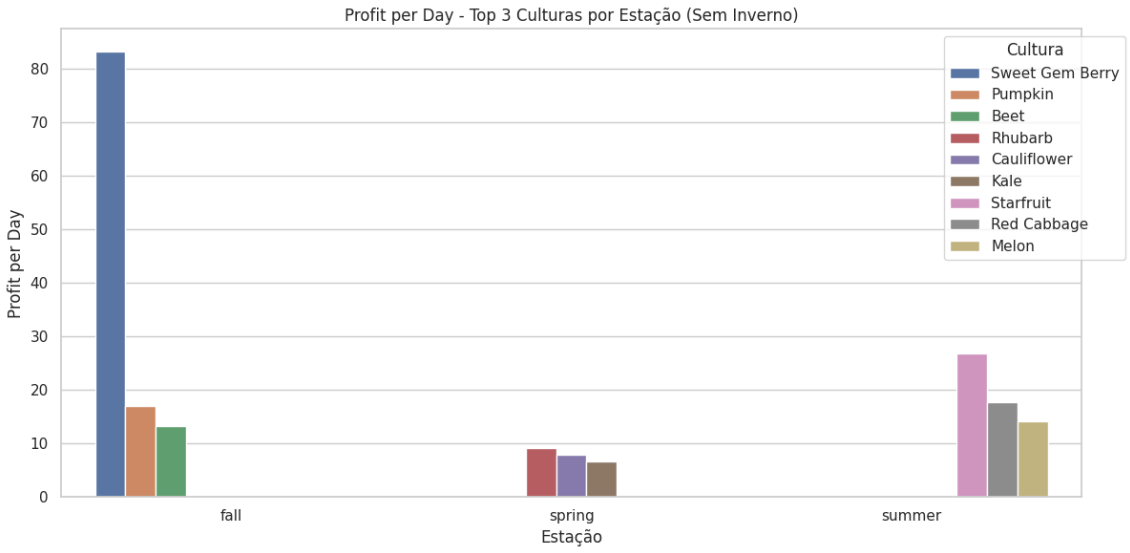
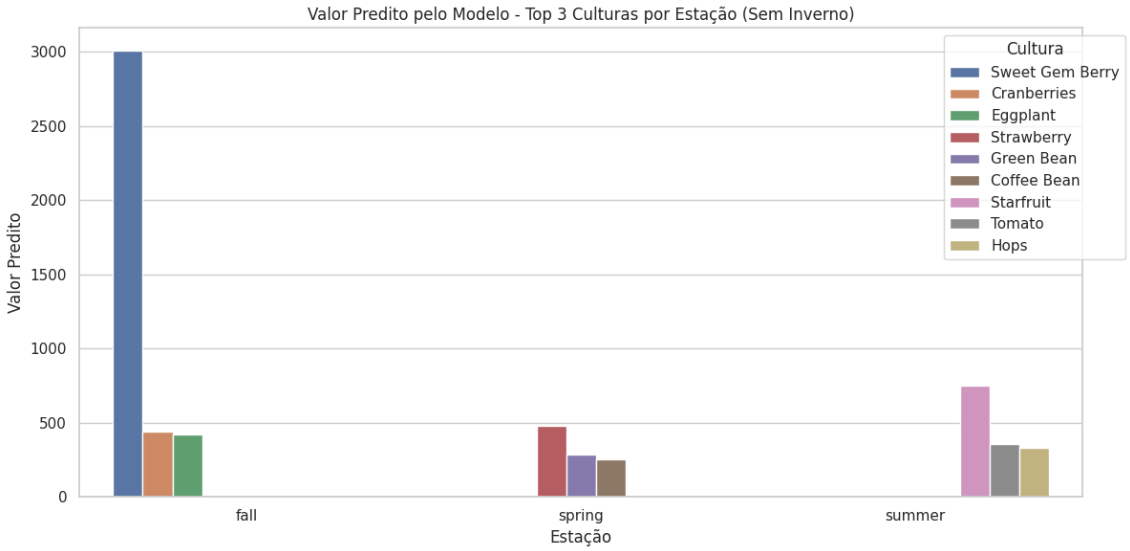


Imagem 32 – O gráfico apresenta as culturas com maior valor estimado pelo modelo ajustado em cada estação, evidenciando as recomendações preditivas para maximização do retorno ao longo do ano produtivo.

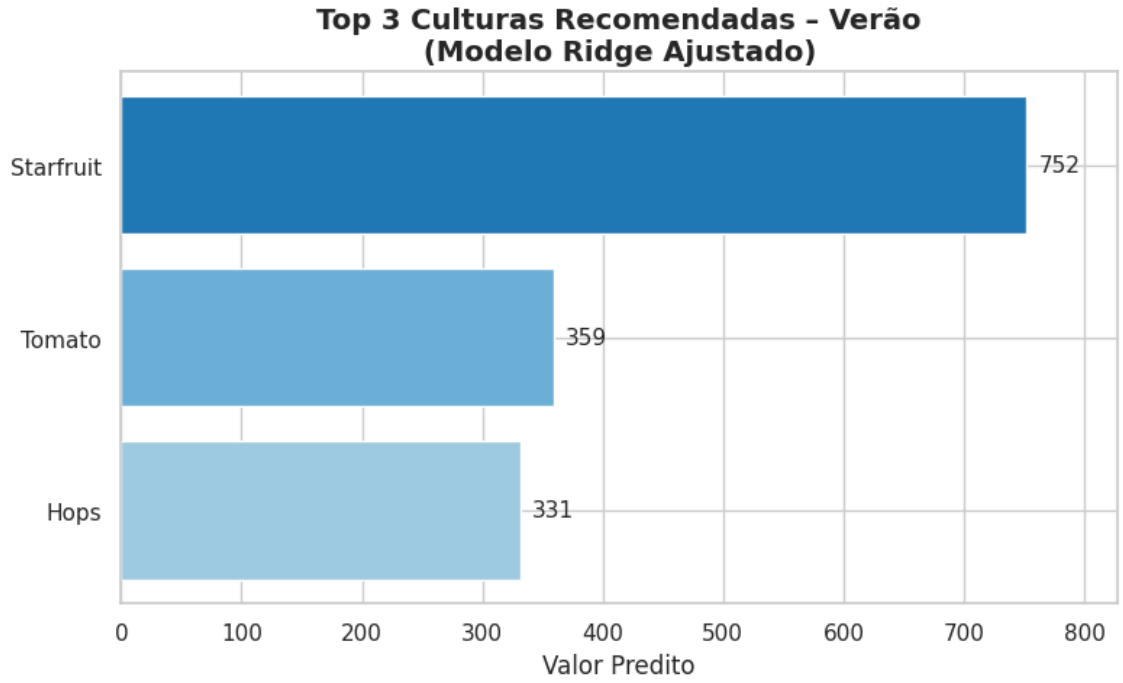


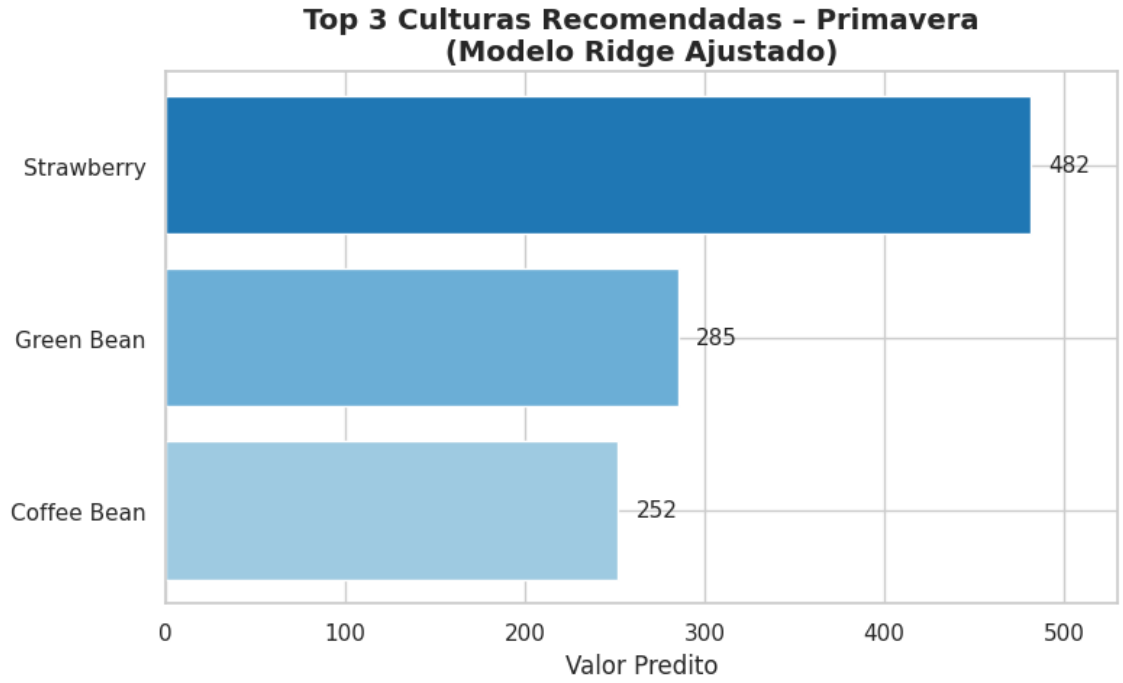
16.1 Qual a melhor estratégia de plantio?

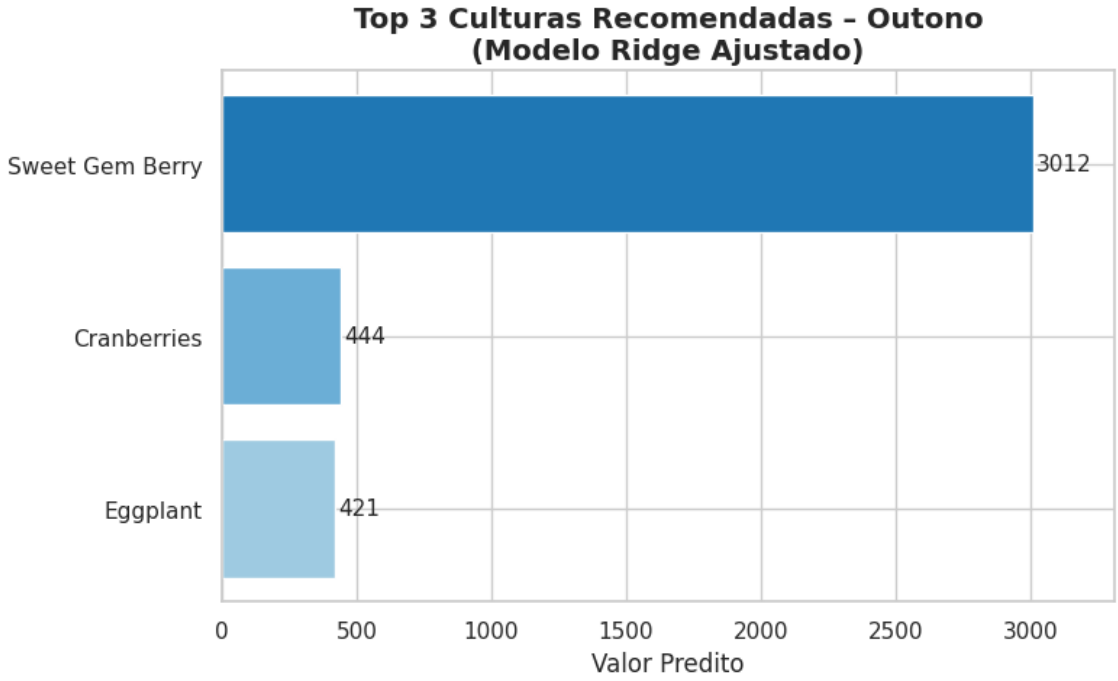
Com base nos resultados do modelo Ridge ajustado (alpha = 10), a estratégia ótima de plantio deve ser definida de forma sazonal, priorizando as culturas com maior valor predito em cada período produtivo. No verão, a Starfruit apresenta o maior potencial de retorno, sendo a principal recomendação, seguida por Tomato e Hops como alternativas complementares. Na primavera, o destaque é o Strawberry, que supera as demais culturas em valor estimado, enquanto Green Bean e Coffee Bean configuram opções secundárias viáveis. Já no outono, a Sweet Gem Berry demonstra desempenho significativamente superior, consolidando-se como a estratégia dominante da estação, com Cranberries e Eggplant como alternativas de suporte.

Assim, a melhor estratégia não consiste em escolher uma única cultura ao longo do ano, mas sim em adaptar o portfólio de plantio conforme a estação, maximizando o retorno previsto pelo modelo e utilizando os recursos de forma mais eficiente em cada ciclo produtivo.

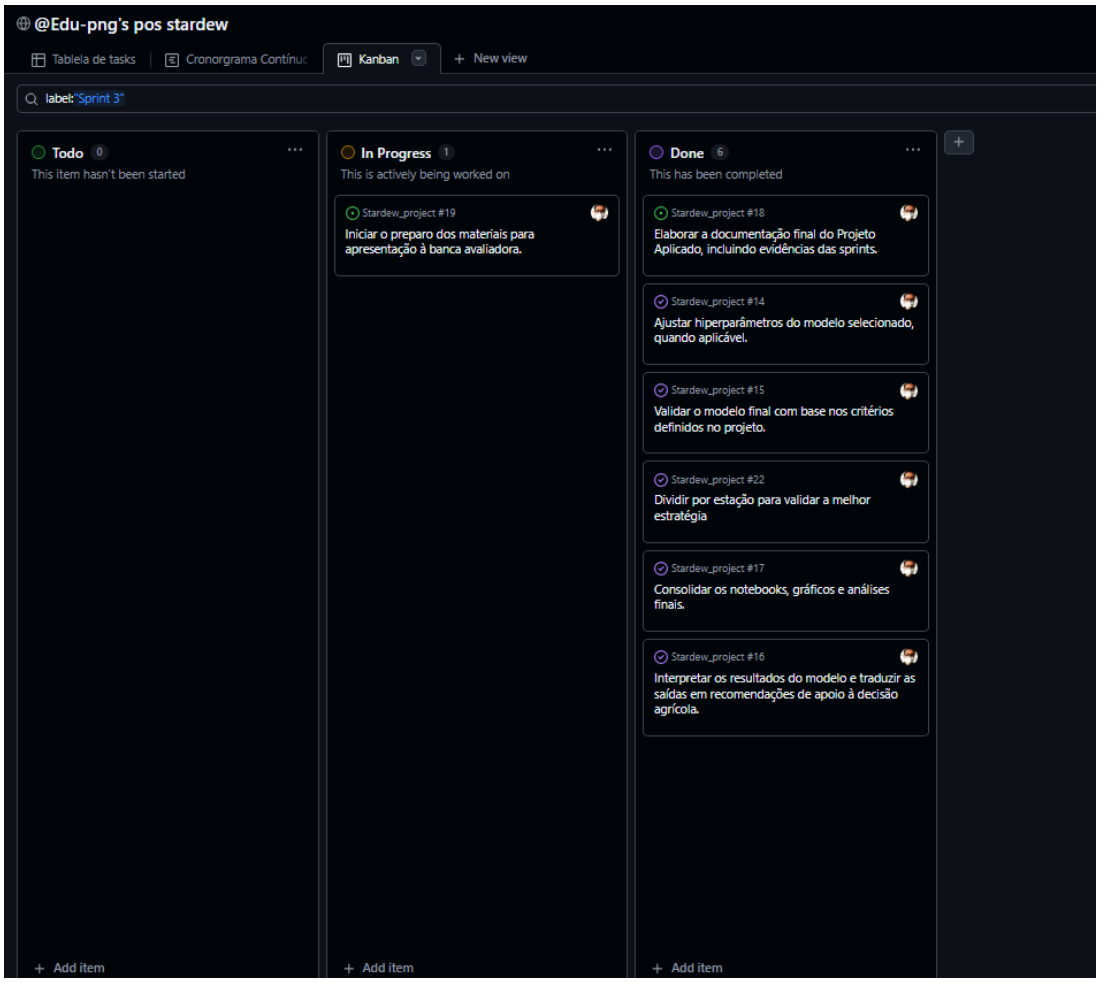
Imagem 29, 28 e 30 - Top 3 Culturas Recomendadas por Estação (Modelo Ridge Ajustado): Os gráficos apresentam as três culturas com maior valor predito pelo modelo para verão, primavera e outono. Observa-se a liderança da Starfruit no verão, Strawberry na primavera e Sweet Gem Berry no outono, evidenciando a recomendação estratégica de plantio otimizada para cada estação produtiva.







- Evidência dos resultados:



A Sprint 3 foi executada conforme planejado, contemplando o ajuste de hiperparâmetros, validação do modelo final, divisão estratégica por estação, consolidação das análicas e elaboração da documentação final do projeto. As atividades concluídas evidenciam a aplicação prática do modelo selecionado para geração de recomendações sazonais de plantio, bem como a interpretação dos resultados em termos de apoio à decisão agrícola. O registro das tarefas no quadro Kanban demonstra organização, rastreabilidade e coerência metodológica na execução da sprint, garantindo transparência e alinhamento com os objetivos definidos para o Projeto Aplicado. A parte de preparo de materiais e gravação do vídeo vai se dar nas semanas seguintes.

2.3.2 Retrospectiva da Sprint

A Sprint 3 foi concluída com êxito, consolidando a etapa de validação final do modelo e sua aplicação prática na geração de recomendações sazonais de plantio. As atividades executadas incluíram o ajuste de hiperparâmetros, análise comparativa de métricas (R^2, MAE e RMSE), divisão estratégica por estação e consolidação dos gráficos e interpretações finais. O uso do quadro Kanban contribuiu para organização, controle de escopo e acompanhamento do progresso, evidenciando clareza na priorização das tarefas e cumprimento dos objetivos planejados.

Como principal aprendizado, destaca-se a importância de alinhar métricas quantitativas com interpretação estratégica, transformando resultados técnicos em recomendações aplicáveis ao contexto agrícola. Além disso, a divisão por estação mostrou-se fundamental para evitar generalizações e garantir decisões mais precisas. A sprint reforçou a relevância da documentação estruturada e da rastreabilidade das decisões de modelagem, assegurando consistência metodológica ao longo do Projeto Aplicado.

3. Considerações Finais

3.1 Resultados

O Projeto Aplicado teve como objetivo desenvolver uma solução analítica capaz de apoiar a decisão de plantio por estação, considerando métricas econômicas como ROI (Retorno sobre Investimento), lucro por dia e valor total estimado ao longo da safra.

A partir da construção de um pipeline de modelagem com regressão Ridge e posterior ajuste de hiperparâmetros (alpha = 10), obteve-se um modelo com desempenho satisfatório, apresentando elevado poder explicativo (R^2 próximo de 0,94) e redução significativa dos erros em relação à configuração inicial. Esse resultado demonstra consistência na capacidade preditiva da solução para estimar o valor econômico das culturas com base nas variáveis disponíveis.

Com a divisão estratégica por estação, foi possível evitar generalizações indevidas e estruturar recomendações específicas para cada período produtivo. Os resultados evidenciam que a melhor estratégia não é fixa ao longo do ano, mas sim dependente do contexto sazonal. No verão, a Starfruit apresentou maior valor predito, indicando alto potencial de retorno. Na primavera, a Strawberry destacou-se como principal recomendação, enquanto no outono a Sweet Gem Berry demonstrou desempenho expressivamente superior às demais culturas.

A análise integrada das métricas reforça que culturas com alto valor absoluto nem sempre possuem o melhor ROI ou lucro diário, tornando essencial a avaliação combinada dos indicadores. Dessa forma, a solução desenvolvida permite não apenas identificar a cultura mais lucrativa, mas também compreender o equilíbrio entre investimento, tempo de crescimento e retorno econômico.

Como ponto positivo, destaca-se a transformação de métricas técnicas em recomendações práticas de plantio, alinhando Data Science à tomada de decisão estratégica. Como limitação, ressalta-se que o cenário analisado considera um ambiente determinístico, sem variáveis externas imprevisíveis (clima real, variação de preços ou riscos agrícolas), o que restringe a complexidade do modelo.

Ainda assim, os resultados alcançados demonstram a viabilidade da abordagem proposta e evidenciam o potencial da modelagem preditiva como ferramenta de apoio à decisão agrícola em contextos estruturados e sazonais. No contexto do jogo, as

estações e cultivos são previsíveis e não se faz necessário um modelo, porém, na realidade os dados apresentam variação de tempo, condições, etc..

Imagem 31 – A figura sintetiza as recomendações finais do modelo ajustado, destacando as três culturas prioritárias para cada estação produtiva. No verão, a Starfruit lidera como principal estratégia, seguida por Tomato e Hops. Na primavera, a Strawberry apresenta o maior potencial, acompanhada por Green Bean e Coffee Bean. No outono, a Sweet Gem Berry se consolida como cultura dominante, com Cranberries e Eggplant como alternativas complementares.

3.2 Próximos passos

Como evolução da solução desenvolvida, recomenda-se a ampliação do modelo para incorporar novas variáveis e restrições, como limite de capital inicial, área disponível para cultivo e combinação simultânea de culturas. A inclusão de validação cruzada e testes com outros algoritmos de regressão também pode aumentar a robustez estatística da solução, permitindo comparar desempenho e reduzir possíveis vieses do modelo atual.

Além disso, um próximo avanço estratégico seria transformar a abordagem em um sistema de simulação de cenários, no qual o usuário possa inserir restrições específicas e receber recomendações personalizadas. Em um contexto real de aplicação agrícola, a incorporação de fatores estocásticos — como variações climáticas e oscilações de mercado — tornaria a solução mais aderente à realidade e ampliaria o papel da modelagem preditiva como ferramenta de apoio à decisão.