Default title text

```
# @title Default title text
%%writefile /content/interim_report.md
# Interim Report: Insurance Risk Analytics for 10 Academ

**Submission Deadline**: June 15, 2025, 8:00 PM UTC
**Repository**: https://github.com/bekonad/insurance-ana
**Prepared by**: Bereket Feleke

This report summarizes progress on **Task 1** (Git and G

## Project Overview
- **Objective**: Analyze historical car insurance data (
- **Data**: `insurance_data.txt`, a pipe-delimited datas
- **Tasks**:
  - Task 1.1: Set up Git and GitHub repository.
  - Task 1.2: Conduct EDA to explore risk and profitabil
  - Task 2: Implement Data Version Control (DVC) for `in
- **Environment**: Mobile device using Google Colab and
- **Tools**: Python, pandas, matplotlib, seaborn, DVC, G

## Task 1: Git and GitHub + Exploratory Data Analysis (E

### Task 1.1: Git and GitHub
- **Status**: Completed
- **Actions**:
  - Created `insurance-analytics` repository with README
  - Configured CI/CD with `.github/workflows/lint.yml` f
  - Merged `task-1` branch into `main` via pull request.
- **Outcome**: Functional repository with version contro

### Task 1.2: EDA
- **Status**: Completed
- **Notebook**: `EDA_Insurance_Analytics.ipynb`
- **Challenges**:
  - **Output Suppression**: Only "Loss Ratio by Gender"
    - **Solution**: Split Loss Ratio code into separate
  - **TypeError**: Temporal trends plot failed due to `P
    - **Solution**: Converted `Month` to `datetime` with
  - **Data Loading**: Initial `KeyError: 'TotalClaims'`
  - **Column Naming**: Renamed columns (`make` to `Make`
  - **Deprecation Warning**: Seaborn's `palette` without

- **EDA Findings**:
- **Loss Ratio by Province**:
 - Gauteng: 0.429, Mpumalanga: 0.393, Limpopo: 0.349, We
 - **Insight**: Gauteng's high loss ratio indicates urba
- **Loss Ratio by VehicleType**:
  - Heavy Commercial: 0.794, Light Commercial: 0.544, Me
  - **Insight**: Heavy Commercial vehicles have the high
- **Loss Ratio by Gender**:
  - Female: 0.492, Male: 0.349, Not specified: 0.348,
```

Default title text ✕ Variables •••

```
# @title Default title text
%%writefile /content/interim_r
# Interim Report: Insurance Ri

**Submission Deadline**: June
**Repository**: https://github
**Prepared by**: Bereket Felek

This report summarizes progres

## Project Overview
- **Objective**: Analyze histo
- **Data**: `insurance_data.tx
- **Tasks**:
  - Task 1.1: Set up Git and G
  - Task 1.2: Conduct EDA to e
  - Task 2: Implement Data Ver
- **Environment**: Mobile devi
- **Tools**: Python, pandas, m

## Task 1: Git and GitHub + Ex

### Task 1.1: Git and GitHub
- **Status**: Completed
- **Actions**:
  - Created `insurance-analyti
  - Configured CI/CD with `.gi
  - Merged `task-1` branch int
- **Outcome**: Functional repo

### Task 1.2: EDA
- **Status**: Completed
- **Notebook**: `EDA_Insurance_
- **Challenges**:
  - **Output Suppression**: On
    - **Solution**: Split Loss
  - **TypeError**: Temporal tr
    - **Solution**: Converted
  - **Data Loading**: Initial
  - **Column Naming**: Renamed
  - **Deprecation Warning**: S

- **EDA Findings**:
- **Loss Ratio by Province**:
 - Gauteng: 0.429, Mpumalanga:
 - **Insight**: Gauteng's high
- **Loss Ratio by VehicleType*
  - Heavy Commercial: 0.794, L
  - **Insight**: Heavy Commerc
- **Loss Ratio by Gender**:
```

```
  - **Insight**: Higher female loss ratio warrants furth
- **Distributions**:
  - `TotalPremium`: Mean 61.91 ZAR, median 2.18 ZAR, max
  - `TotalClaims`: Mean 64.86 ZAR, median 0 ZAR, right-s
  - **Insight**: Most policies have low premiums and zer
- **Temporal Trends**:
  - Average claims and claim frequency plotted, showing
  - **Insight**: Likely seasonal patterns (e.g., Q4 spik
- **Claims by Make**:
  - Analysis intended but not shown in output.
  - **Insight**: Expected high claims for prevalent make

 - **Outcome**: Comprehensive EDA addressing all guiding

## Task 2: Data Version Control (DVC)
- **Status**: Completed
- **Notebook**: `DVC_Setup.ipynb`
- **Actions**:
  - Initialized DVC with `--no-scm` to resolve SCM error
  - Configured local storage at `/content/dvc_storage`.
  - Tracked `insurance_data.txt` with `dvc add` and push
  - Committed `insurance_data.txt.dvc` and `.dvc/config`
  - Verified with `dvc pull`.
- **Challenges**:
  - SCM error resolved with `--no-scm`.
  - Mobile constraints addressed using local storage and
- **Outcome**: Dataset tracked, meeting data versioning
##Next Steps
- Merge `task-2` pull request.
- Start Task 3 (A/B Testing) using EDA insights.
- Submit repository URL, PDF of `EDA_Insurance_Analytics

## Citations
- [DVC Getting Started](https://dvc.org/doc/start)
- [Google Colab](https://colab.research.google.com/noteb
- [GitHub Docs](https://docs.github.com/en/repositories/

Overwriting /content/interim_report.md
```

```
  - Female: 0.492, Male: 0.349
  - **Insight**: Higher female
- **Distributions**:
  - `TotalPremium`: Mean 61.91
  - `TotalClaims`: Mean 64.86
  - **Insight**: Most policies
- **Temporal Trends**:
  - Average claims and claim f
  - **Insight**: Likely season
- **Claims by Make**:
  - Analysis intended but not
  - **Insight**: Expected high

 - **Outcome**: Comprehensive

## Task 2: Data Version Contro
- **Status**: Completed
- **Notebook**: `DVC_Setup.ipy
- **Actions**:
  - Initialized DVC with `--no
  - Configured local storage a
  - Tracked `insurance_data.tx
  - Committed `insurance_data.
  - Verified with `dvc pull`.
- **Challenges**:
  - SCM error resolved with `-
  - Mobile constraints address
- **Outcome**: Dataset tracked
##Next Steps
- Merge `task-2` pull request.
- Start Task 3 (A/B Testing) u
- Submit repository URL, PDF o

## Citations
- [DVC Getting Started](https:
- [Google Colab](https://colab
- [GitHub Docs](https://docs.g

Overwriting /content/interim_rep
```