

```
!git clone https://github.com/bekonad/insurance-analytics.git
%cd /content/insurance-analytics
```



```
Cloning into 'insurance-analytics'...
remote: Enumerating objects: 54, done.
remote: Counting objects: 100% (54/54), done.
remote: Compressing objects: 100% (47/47), done.
remote: Total 54 (delta 14), reused 0 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (54/54), 611.65 KiB | 11.33 MiB/s, done.
Resolving deltas: 100% (14/14), done.
/content/insurance-analytics
```

```
!pip install pandas numpy matplotlib seaborn scipy dvc
```



```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (2.0.2
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (
Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.1
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (1.15.
Collecting dvc
  Downloading dvc-3.60.1-py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dis
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: cycycler>=0.10 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (f
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-package
Collecting celery (from dvc)
  Downloading celery-5.5.3-py3-none-any.whl.metadata (22 kB)
Collecting colorama>=0.3.9 (from dvc)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Collecting configobj>=5.0.9 (from dvc)
  Downloading configobj-5.0.9-py2.py3-none-any.whl.metadata (3.2 kB)
Requirement already satisfied: distro>=1.3 in /usr/local/lib/python3.11/dist-packages
Collecting dpath<3,>=2.1.0 (from dvc)
  Downloading dpath-2.2.0-py3-none-any.whl.metadata (15 kB)
Collecting dulwich (from dvc)
  Downloading dulwich-0.22.8-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.wh
Collecting dvc-data<3.17,>=3.16.2 (from dvc)
  Downloading dvc_data-3.16.10-py3-none-any.whl.metadata (5.0 kB)
Collecting dvc-http>=2.29.0 (from dvc)
  Downloading dvc_http-2.32.0-py3-none-any.whl.metadata (1.3 kB)
Collecting dvc-objects (from dvc)
  Downloading dvc_objects-5.1.1-py3-none-any.whl.metadata (3.8 kB)
Collecting dvc-render<2,>=1.0.1 (from dvc)
  Downloading dvc_render-1.0.2-py3-none-any.whl.metadata (5.4 kB)
Collecting dvc-studio-client<1,>=0.21 (from dvc)
  Downloading dvc_studio_client-0.21.0-py3-none-any.whl.metadata (4.3 kB)
```

```

Collecting dvc-task<1,>=0.3.0 (from dvc)
  Downloading dvc_task-0.40.2-py3-none-any.whl.metadata (10.0 kB)
Collecting flatten_dict<1,>=0.4.1 (from dvc)
  Downloading flatten_dict-0.4.2-py2.py3-none-any.whl.metadata (9.2 kB)
Collecting fluf1.lock<9,>=8.1.0 (from dvc)
  Downloading fluf1_lock-8.2.0-py3-none-any.whl.metadata (3.3 kB)
Requirement already satisfied: fsspec>=2024.2.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Collecting funcy>=1.14 (from dvc)
  Downloading funcy-2.0-py2.py3-none-any.whl.metadata (5.9 kB)
Collecting grandalf<1,>=0.7 (from dvc)
  Downloading grandalf-0.8-py3-none-any.whl.metadata (1.7 kB)
Collecting gto<2,>=1.6.0 (from dvc)
  Downloading gto-1.7.2-py3-none-any.whl.metadata (4.7 kB)
Collecting hydra-core>=1.1 (from dvc)
  Downloading hydra_core-1.3.2-py3-none-any.whl.metadata (5.5 kB)
Collecting iterative-telemetry>=0.0.7 (from dvc)
  Downloading iterative_telemetry-0.0.10-py3-none-any.whl.metadata (4.1 kB)
Collecting kombu (from dvc)

```

!pip install dvc



```

Requirement already satisfied: dvc in /usr/local/lib/python3.11/dist-packages (3.60.1)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: celery in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: colorama>=0.3.9 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: configobj>=5.0.9 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: distro>=1.3 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dpath<3,>=2.1.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dulwich in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-data<3.17,>=3.16.2 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-http>=2.29.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-objects in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-render<2,>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-studio-client<1,>=0.21 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: dvc-task<1,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: flatten_dict<1,>=0.4.1 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: fluf1.lock<9,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: fsspec>=2024.2.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: funcy>=1.14 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: grandalf<1,>=0.7 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: gto<2,>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: hydra-core>=1.1 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: iterative-telemetry>=0.0.7 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: kombu in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: networkx>=2.5 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: omegaconf in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: packaging>=19 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: pathspec>=0.10.3 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: platformdirs<5,>=3.1.1 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: psutil>=5.8 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: pydot>=1.2.4 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: pytrie>=2.3.2 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: pyparsing>=2.4.7 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: requests>=2.22 in /usr/local/lib/python3.11/dist-packages (from dvc)
Requirement already satisfied: rich>=12 in /usr/local/lib/python3.11/dist-packages (from dvc)

```

```
Requirement already satisfied: ruamel.yaml>=0.17.11 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: scmrepo<4,>=3.3.8 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: shortuuid>=0.5 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: shtab<2,>=1.3.4 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: tabulate>=0.8.7 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: tomlkit>=0.11.1 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: tqdm<5,>=4.63.1 in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: voluptuous>=0.11.7 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: zc.lockfile>=1.2.1 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: dictdiffer>=0.8.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: diskcache>=5.2.1 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: sqltrie<1,>=0.11.0 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: orjson<4,>=3 in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: aiohttp-retry>=2.5.0 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: billiard<5.0,>=4.2.1 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: vine<6.0,>=5.1.0 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: click<9.0,>=8.1.2 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: click-didyoumean>=0.3.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: click-repl>=0.2.0 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: click-plugins>=1.1.1 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dis
Requirement already satisfied: six<2.0,>=1.12 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: atpublic in /usr/local/lib/python3.11/dist-packages (fr
```

```
!dvc --version
```

```
➡ 3.60.1
```

```
!dvc init
```

```
➡ ERROR: failed to initiate DVC - '.dvc' exists. Use `-f` to force.
```

```
!cat .dvc/config
```

```
!dvc pull
```

```
➡ Collecting          |0.00 [00:00,   ?entry/s]
   Fetching
   Building workspace index          |1.00 [00:01, 1.15s/entry]
   Comparing indexes          |2.00 [00:00, 1.13kentry/s]
   Applying changes          |0.00 [00:00,   ?file/s]
   No remote provided and no default remote set.
   Everything is up to date.
```

```
!ls -la
```

```
➡ total 517300
drwxr-xr-x 6 root root    4096 Jun 18 22:54 .
drwxr-xr-x 1 root root    4096 Jun 18 22:54 ..
-rw-r--r-- 1 root root     108 Jun 18 20:32 config
drwxr-xr-x 3 root root    4096 Jun 18 20:32 docs
```

```

drwxr-xr-x 3 root root      4096 Jun 18 20:33 .dvc
-rw-r--r-- 1 root root        139 Jun 18 20:33 .dvcignore
-rw-r--r-- 1 root root        139 Jun 18 20:32 dvcignore
-rw-r--r-- 1 root root    290271 Jun 18 20:32 EDA_Insurance_Analytics.ipynb
drwxr-xr-x 8 root root      4096 Jun 18 20:33 .git
drwxr-xr-x 3 root root      4096 Jun 18 20:32 .github
-rw-r--r-- 1 root root     4319 Jun 18 20:32 .gitignore
-rw-r--r-- 1 root root 529363713 Jun 18 22:18 insurance_data.txt
-rw-r--r-- 1 root root        103 Jun 18 20:32 insurance_data.txt.dvc
-rw-r--r-- 1 root root       1509 Jun 18 20:32 README.md

```

```

import pandas as pd
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import os

# Load and prepare data
file_path = 'insurance_data.txt'
if not os.path.exists(file_path):
    raise FileNotFoundError(f"{file_path} not found. Please upload the file.")
data = pd.read_csv(file_path, sep='|')
data = data.dropna(subset=['TotalPremium', 'TotalClaims', 'Province', 'VehicleType', 'Profit'])
data['Profit'] = data['TotalPremium'] - data['TotalClaims']

# Define low-risk segments based on EDA insights
low_risk_segments = ['Free State', 'Bus', 'Male']
treatment_group = data[
    (data['Province'].isin(['Free State'])) |
    (data['VehicleType'].isin(['Bus'])) |
    (data['Gender'].isin(['Male']))
]
control_group = data[
    ~(data['Province'].isin(['Free State'])) |
    data['VehicleType'].isin(['Bus']) |
    data['Gender'].isin(['Male'])
]

# Sample 50% for A/B test
treatment_sample = treatment_group.sample(frac=0.5, random_state=42)
control_sample = control_group.sample(frac=0.5, random_state=42)

# Apply 10% premium reduction to treatment group
treatment_sample['NewPremium'] = treatment_sample['TotalPremium'] * 0.90
treatment_sample['NewProfit'] = treatment_sample['NewPremium'] - treatment_sample['TotalClaims']

# Control group remains unchanged
control_sample['NewProfit'] = control_sample['Profit']

# Visualize results
plt.figure(figsize=(10, 6))
plt.boxplot([treatment_sample['NewProfit'].dropna(), control_sample['NewProfit'].dropna()])

```

```

labels=['Treatment (10% Reduction)', 'Control (No Change)'])
plt.title('Profitability Comparison: A/B Test')
plt.ylabel('Profit (ZAR)')
plt.show()

# Perform t-test
t_stat, p_value = stats.ttest_ind(treatment_sample['NewProfit'].dropna(),
                                   control_sample['NewProfit'].dropna(), equal_v

print(f"T-statistic: {t_stat:.4f}")
print(f"P-value: {p_value:.4f}")

```

```

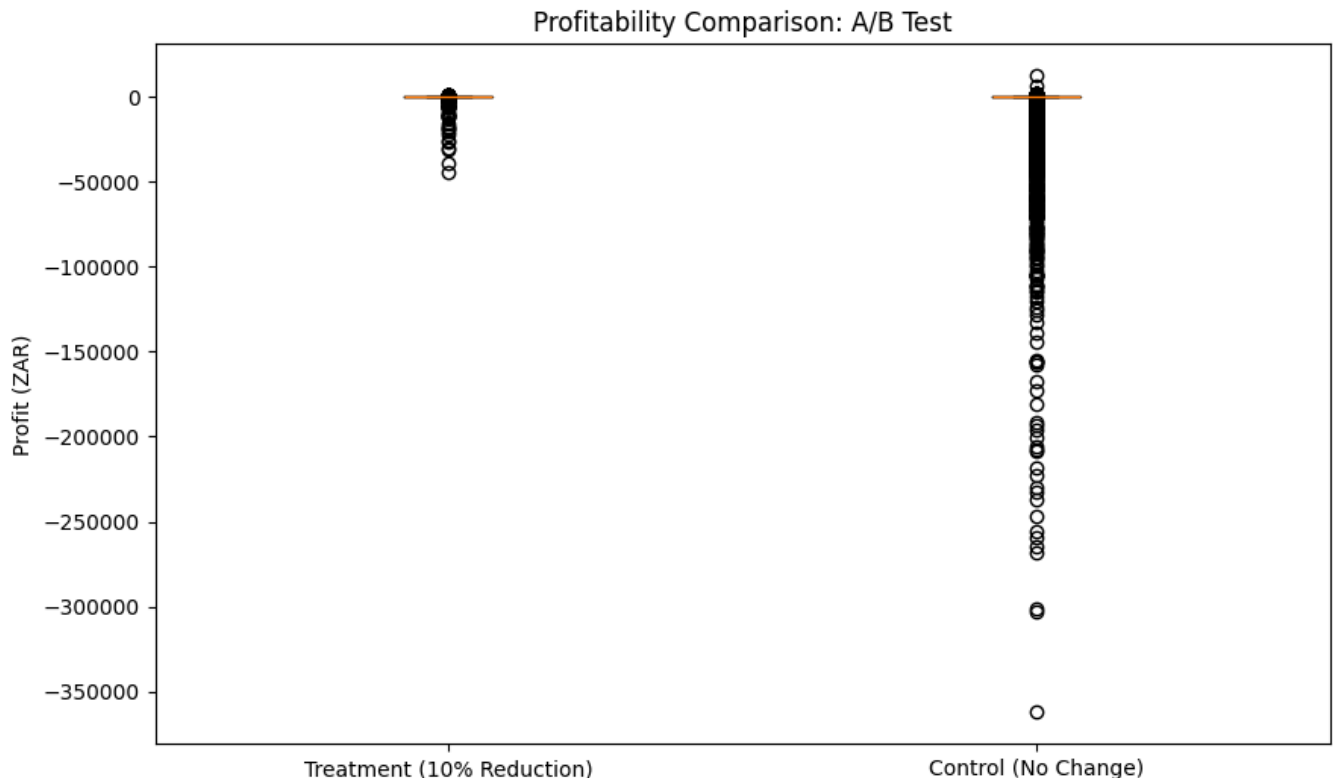
# Interpret results
if p_value < 0.05:
    print("Conclusion: There is a significant difference in profitability with
else:
    print("Conclusion: No significant difference in profitability with the 10%

```

➡ /tmp/ipython-input-24-1166075729.py:11: DtypeWarning: Columns (32,37) have mixed types.
data = pd.read_csv(file_path, sep='|')
/tmp/ipython-input-24-1166075729.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user>
data['Profit'] = data['TotalPremium'] - data['TotalClaims']

/tmp/ipython-input-24-1166075729.py:41: MatplotlibDeprecationWarning: The 'labels' param
plt.boxplot([treatment_sample['NewProfit'].dropna(), control_sample['NewProfit'].dropna()])



T-statistic: 3.9374

P-value: 0.0001

Conclusion: There is a significant difference in profitability with the 10% premium reduction.

