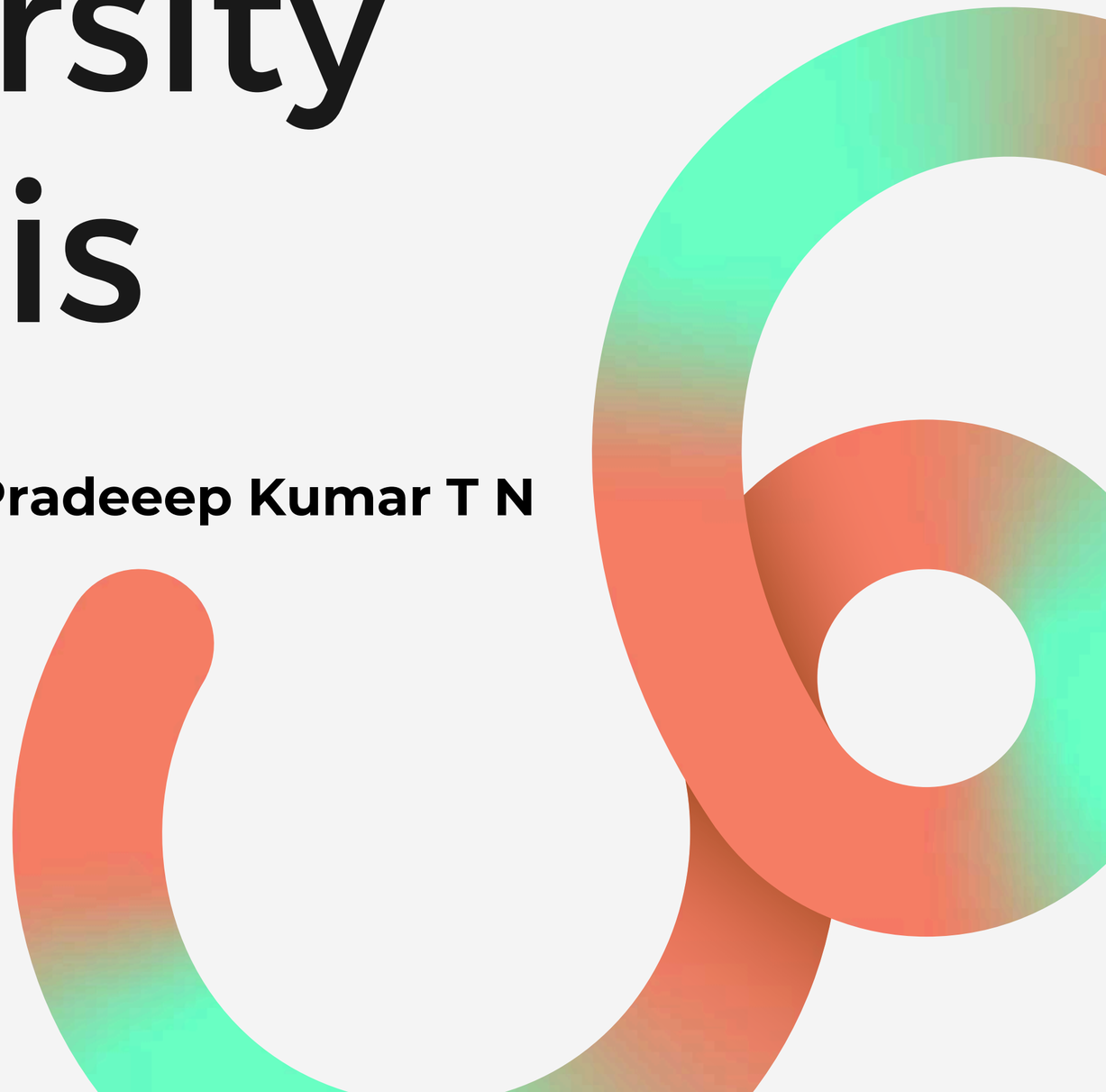
Open University Data Analysis

By

Pradeeep Kumar T N



Topic Outlines

- Course_Intake
- Course_Module
- Gender
- Region
- Highest Qualification
- IMD Band
- Age Band
- No of Previous Attempts
- Studied Credits
- Disability
- Final Result

Steps

1. Define Objectives
2. Collect Data
3. Data Wrap Up
4. Data Cleaning
5. Data Exploration
6. Data Transformation
7. Interpret Results
8. Statistical Analysis

Data Overview

The Open University (OU) in the UK is a distinguished institution known for its focus on distance learning and flexible education.

- Established: 1969
- Type: Public distance learning and research university
- Location: Milton Keynes, Buckinghamshire, UK
- Mission: To provide accessible, flexible, and high-quality education to students from diverse backgrounds.

Importing Packages

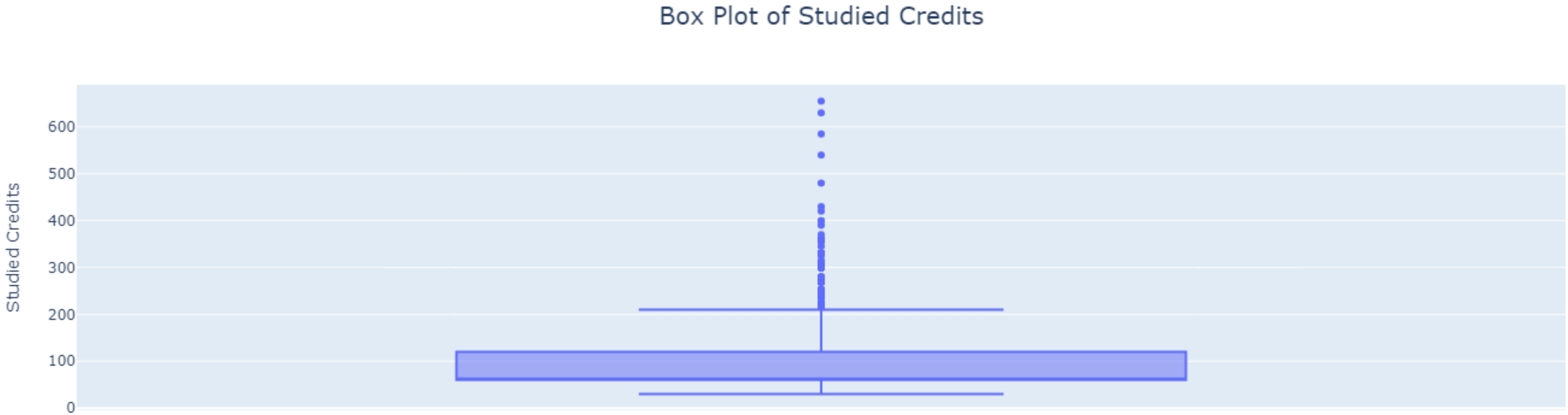
Python Packages

- import numpy as np
- import pandas as pd
- import matplotlib.pyplot as plt
- import plotly.express as pt
- import plotly.graph_objects as go
- import seaborn as sns
- from sklearn.preprocessing import OneHotEncoder
- import os
- import warnings
- warnings.filterwarnings("ignore")

Statistical Packages

- import scipy.stats as stats
- import statsmodels.api as sm
- from scipy.stats import ttest_1samp
- from scipy.stats import ttest_ind
- from scipy.stats import ttest_rel
- from scipy.stats import f_oneway
- from statsmodels.formula.api import ols

Checking and Removing Outliers

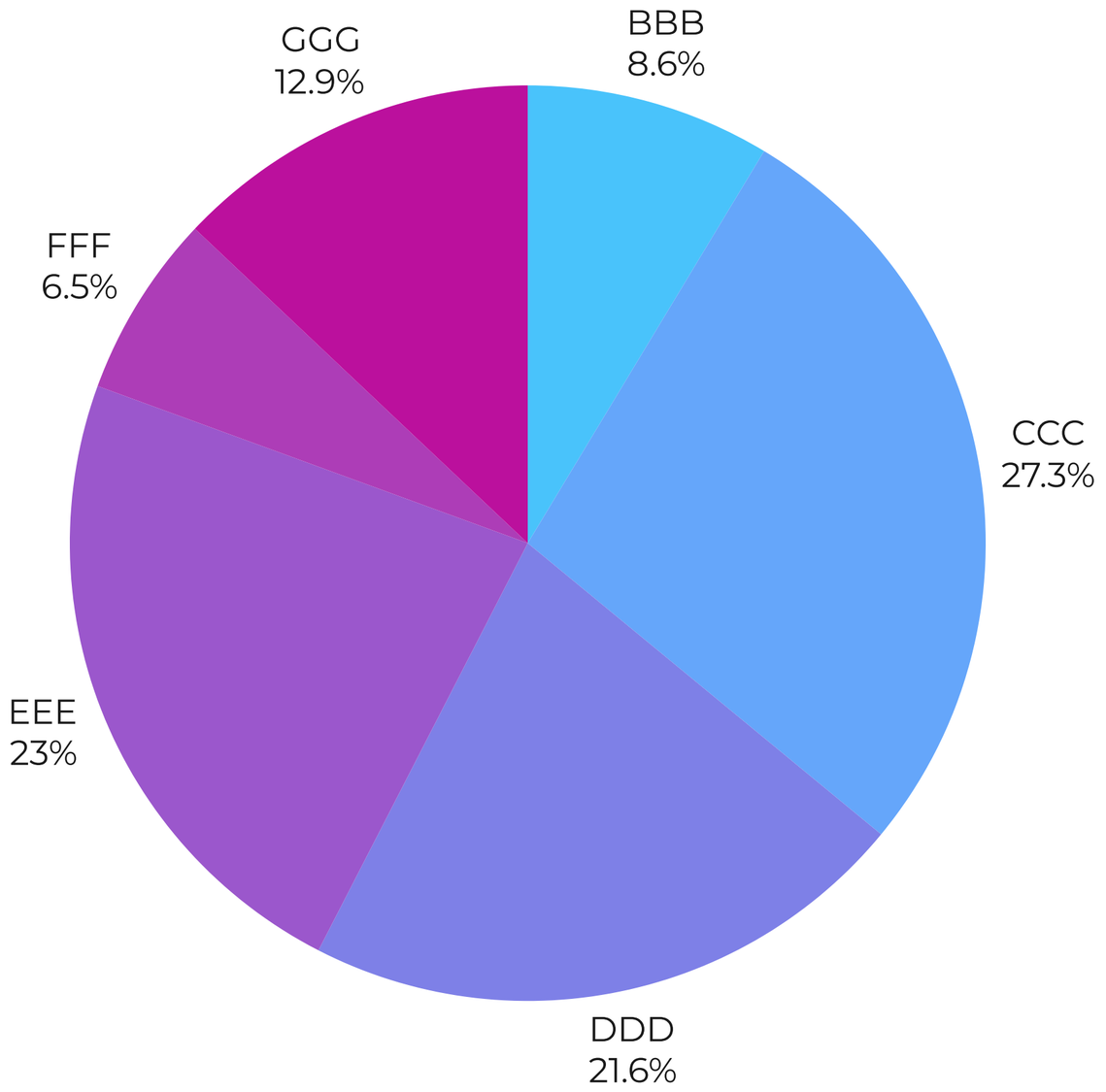


studied_credits lower_bound : -30.0
studied_credits upper_bound : 210.0
studied_credits median_value : 60.0

Pie Charts

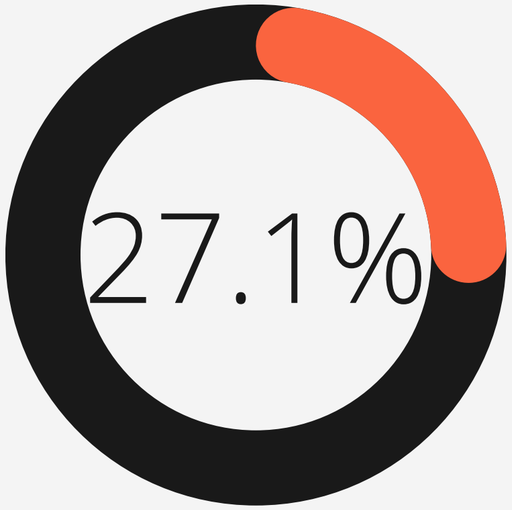
Course_Module

- Grouped Pie Charts: Place the pie charts side by side for direct visual comparison.

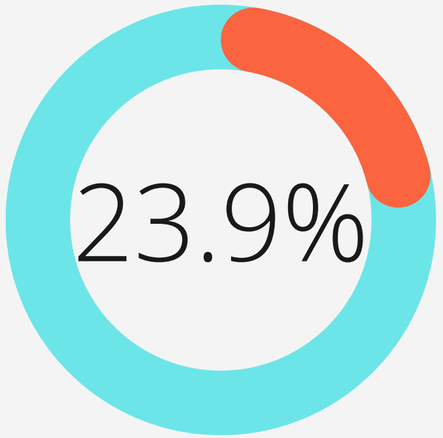


Donut Chart

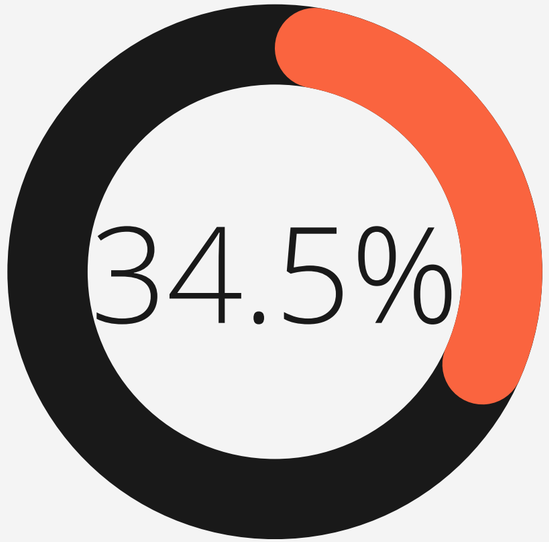
2013-Feb
Course_Intake



2013-Oct
Course_Intake



2014-Feb
Course_Intake

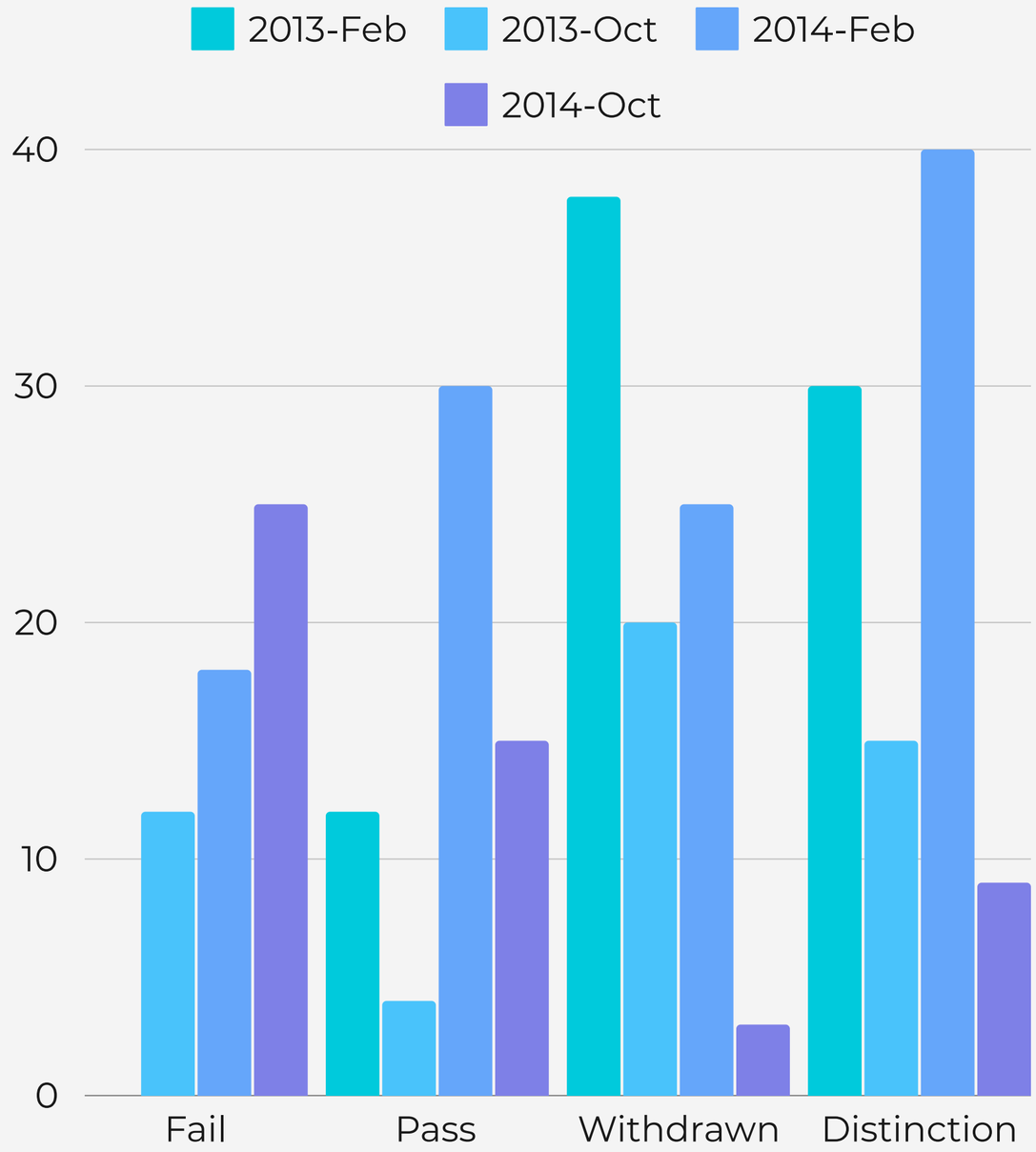


2014-Oct
Course_Intake

Bar Charts

Course_Intake vs Final Result

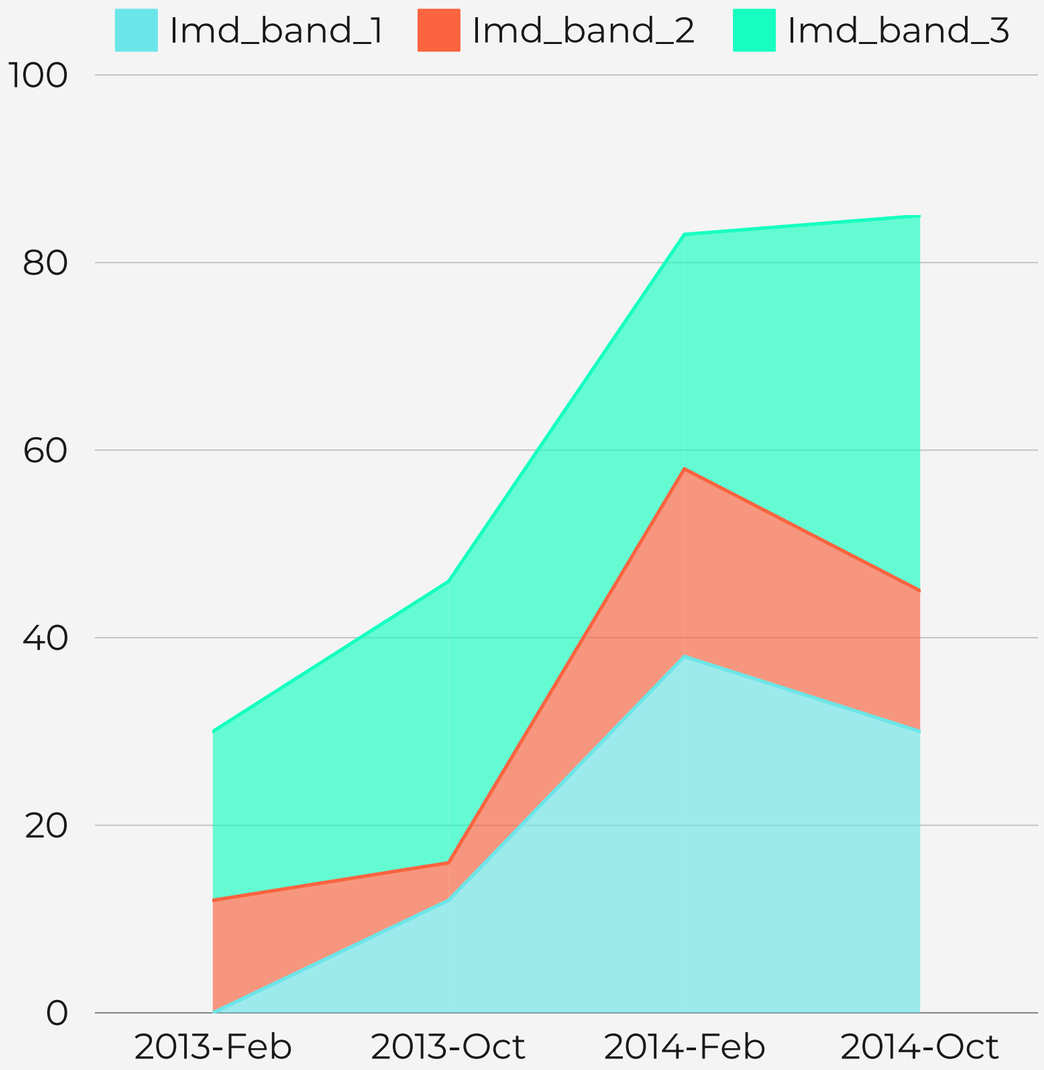
- Summarize the overall trends and their potential impact.



Trend Report

Course_Intake vs IMD_Band

- Recap the main trends and findings from the analysis.
- Provide an overview of how deprivation levels have changed and what it means for the affected areas.



Statistical Analysis

ttest_ind Statistic: 4.767091609279027
T-statistic: 4.767091609279027
P-value: 1.8771022866691152e-06
Reject the null hypothesis at alpha = 0.05. There is a significant difference in studied credits between males and females.

Chi-square Statistic: 4.0
P-value: 0.26146412994911117
Degrees of Freedom: 3
Expected Frequencies Table:
[[0.5 0.5 0.5 0.5]
[0.5 0.5 0.5 0.5]]
Fail to reject the null hypothesis at alpha = 0.05. There is no significant association between gender and course intake.

Thank you

Email

pradeepkumartn17@gmail.com

Social Media

@linkedin

By

Pradeep Kumar T N