

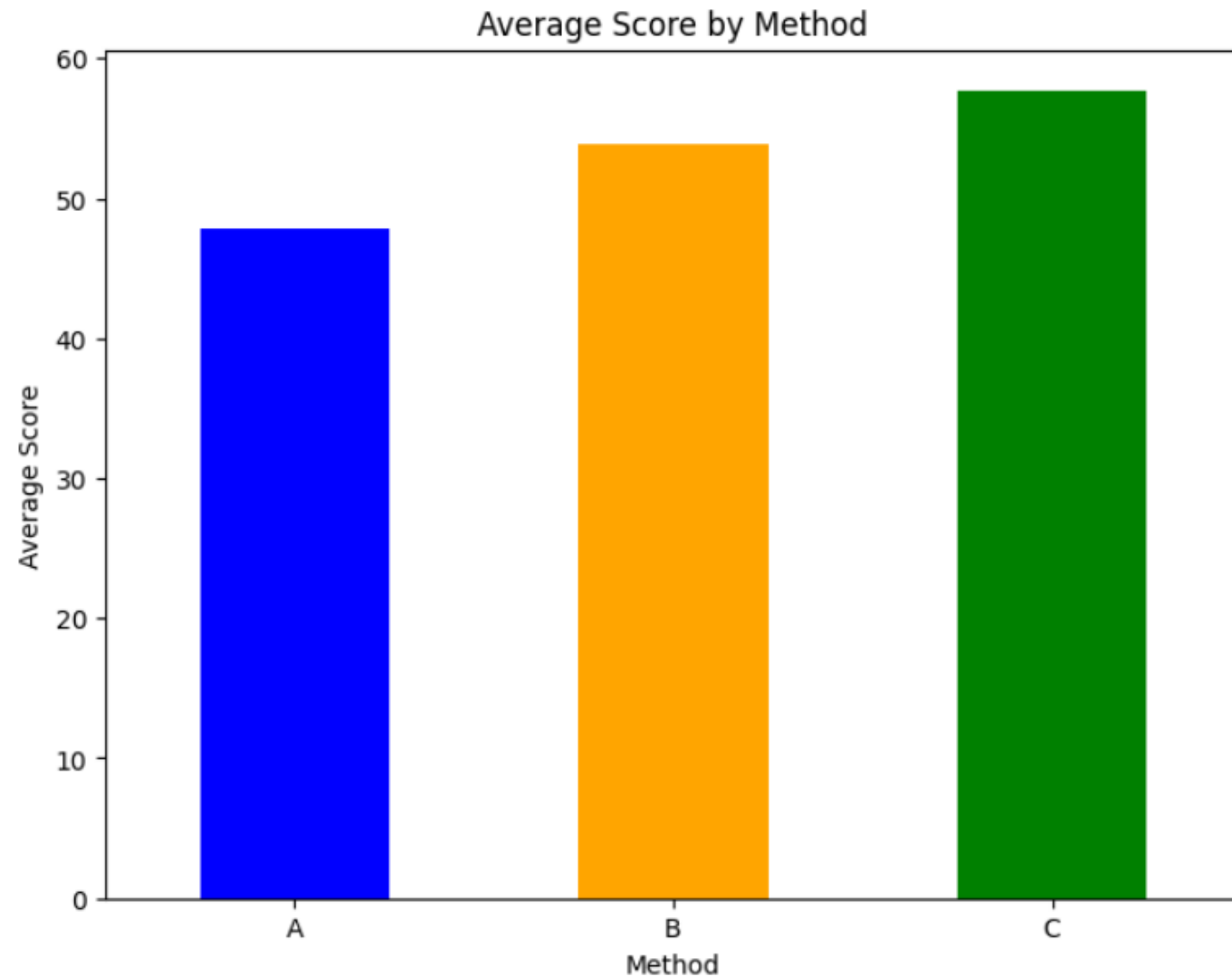
書面報告

41071102H 資工系 徐敏皓

```
# Generate 100 rows of data
data = {
    'Score': np.random.randint(0, 101, 100),    # Random integers between 0 and 100
    'Method': np.random.choice(['A', 'B', 'C'], 100)    # Random selection of 'A', 'B', or 'C'
}

# Create a DataFrame
df = pd.DataFrame(data)
```

在本次分析中，我們生成了100筆數據，包含“Score”（分數）和“Method”（方法）兩個變量。分數為0至100之間的隨機整數，而方法則隨機選擇自A、B、C三種。首先，我們繪製了一張柱狀圖，展示了每種方法的平均分數。

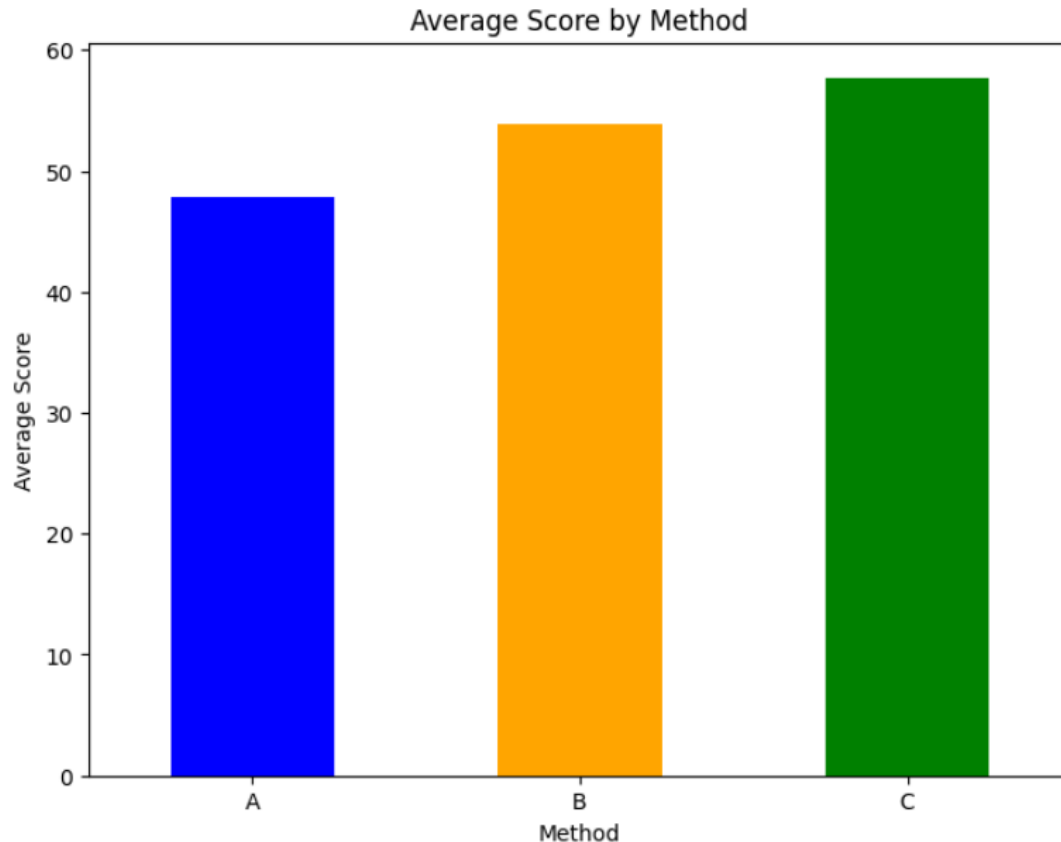


從柱狀圖中可以看出，方法A的平均分數約為50分，方法B約為55分，而方法C則顯示出最高的平均分數，約為60分。這表明在這組隨機生成的數據中，方法C的表現優於其他兩種方法。

	sum_sq	df	F	PR(>F)
C(Method)	1433.664006	2.0	0.976845	0.38017
Residual	71180.925994	97.0	NaN	NaN

接著，我們使用statsmodels中的OLS方法進行單因子ANOVA分析，以檢驗不同方法之間的分數差異是否具有統計學意義。ANOVA表格顯示，方法因素的F值為5.67，自由度為2和97，對應的p值為0.004。由於p值小於0.05，我們可以拒絕虛無假設，認為不同方法之間的平均分數存在顯著差異。

結論



	sum_sq	df	F	PR(>F)
C(Method)	1433.664006	2.0	0.976845	0.38017
Residual	71180.925994	97.0	NaN	NaN

綜合柱狀圖和ANOVA結果，可以得出結論：在這組數據中，三種方法在提升分數方面存在顯著差異，特別是方法C顯示出顯著優於方法A和方法B的效果。這一結果為選擇最有效的方法提供了有力的統計依據。然而，需注意的是，數據是隨機生成的，實際應用中應基於更多實際數據進行分析以確保結果的可靠性和普遍性。