1. Data Summary
serial no.: 申請者的編號，從 1 到 500 號，共 500 人。
GRE. Score: GRE成績，分為 verbal 和 quantitative 兩部份，各 170 分，本資料中最低分 290 分，最高分 340 分，平均 316 分。
TOEFL. Score: 托福成績，分為聽說讀寫四科，每科各 30 分，本資料中最低分 92 分，最高分 120 分，平均 107。
University. Rating: 大學排名，分為五個等級，最低等為 1，最高等為 5。
SOP: Statement of purpose 申請動機，最低 1 分，最高 5 分。
LOR: Letter of recommendation 推薦信強度，最低 1 分，最高 5 分。
CGPA: Undergraduate GPA在校成績，最高分 10 分，本資料中最高 9.92，最低 6.8，平均 8.58。
Research: 申請者是否發表研究論文或相關著作，未發表為 0，有發表為 1。
Chance. of. Admit: 申請者錄取 UCLA 的機率。

2. Bar plot
GRE Score：GRE的平均數為 316.68 分。從圖中可以看出在310~315區間以及320~325區間有一個高峰，顯示絕大多數學生的分數落在320分附近。
TOEFL Score：TOEFL的平均分數為107.19分。大多數學生的成績位於100~110分的區間內，其中又以考取105~110分的學生人數為最多。
SOP：資料將 Statement of Purpose 分為五個級距，大多數申請人的 SOP 落在 3~4 的級距內。
LOR：資料將 Letter of Recommendation 分為五個級距，大多數申請人的LOR落在 3~4 的級距內。
CGPA：資料將大學的GPA分為 10 個級距，申請人最低的 CGPA 落在等級 7 的地方，而大多數的人則是落在 8~9 的級距內。
University Rating：資料將 University Rating 分為 5 個級距，大多數的申請人來自於排名位於中段的學校。只有極少數的人來自排名位於後段的學校。

3. 從 Correlation plot 可以看出 GRE、TOEFL 成績以及大學 GPA 與錄取率有很大的相關性。
GRE、TOEFL 成績與錄取率的相關係數約為 0.8，大學 GPA 與錄取率的相關係數則將近 0.9。而 Research 與錄取率的相關係數只有 0.55，表示相對於其他條件，申請人是否具有研究經驗可能不是那麼的重要。另外 GRE、TOEFL 成績以及大學 GPA 兩倆之間的相關程度也很高，表示申請人在大學學業成績的表現也反映在入學考試以及語言成績上。

4. Data Preprocessing
對各個變數做 MinMax Scaling，以降低 Data Domain Range 的影響。

$$\frac{X_i - min(X)}{X_{max} - X_{min}}$$

7. Clustering:
希望能夠依照所得特徵來來將學生生作分群，並得出結論
一、找出最好的群數量
1.SSE - 2
2.Silhouette - 2
3.Gap stat - 5
4.最後選擇 4 來來當群的數量量
二、用不同的方法去分群

1. simple kmeans
2. dist cluster method

三、分群所得圖，群聚關係並不明顯，因此試著挑一一些特徵出去，讓分群效果更清楚
1.AIC - Forward & Backward
2.BIC - Forward & Backward
3.最重要的特徵為 GRE、GPA、TOFEL、LOR，其中以 GRE 跟 GPA 所得之圖分群效果最好