**認識資料集**

Data Set Characteristics: Multivariate　所以我們用來分析或是資料視覺化是很合理的

此資料為2013年蒐集的

從 Gazi University in Ankara (Turkey).

There is a

* total of 28 course specific questions
* additional 5 attributes.

號稱 Missing Values? N/A

認識資料

Attribute Information: instr: Instructor's identifier; values taken from {1,2,3}

教職人員的編號 共3人

class: Course code (descriptor); values taken from {1-13}

課程代碼 共 1~13種課程

repeat: Number of times the student is taking this course; values taken from {0,1,2,3,...}

學生學習的次數

attendance: Code of the level of attendance; values from {0, 1, 2, 3, 4}

出勤率

difficulty: Level of difficulty of the course as perceived by the student; values taken from {1,2,3,4,5}

學生認為的難度

**問卷內容(不用全放入PPT，挑你覺得相關的或我再處理，Q1~Q28會額外印下來提供給老師)**

Q1: The semester course content, teaching method and evaluation system were provided at the start.

一开始就提供了本学期的课程内容、教学方法和评价体系。

Q2: The course aims and objectives were clearly stated at the beginning of the period.

课程的目的和目标在开始时就已经明确说明。

Q3: The course was worth the amount of credit assigned to it.

该课程的价值与分配给它的学分相同。

Q4: The course was taught according to the syllabus announced on the first day of class.

课程是按照第一天上课时公布的教学大纲进行的。

Q5: The class discussions, homework assignments, applications and studies were satisfactory.

课堂讨论、家庭作业、应用和学习都令人满意。

Q6: The textbook and other courses resources were sufficient and up to date.

教科书和其他课程资源是充分和最新的。

Q7: The course allowed field work, applications, laboratory, discussion and other studies.

该课程允许实地工作、应用、实验室、讨论和其他研究。

Q8: The quizzes, assignments, projects and exams contributed to helping the learning.

测验、作业、项目和考试对学习有帮助。

Q9: I greatly enjoyed the class and was eager to actively participate during the lectures.

我非常喜欢这门课，并渴望在讲座中积极参与。

Q10: My initial expectations about the course were met at the end of the period or year.

我最初对这门课的期望在期末或年终得到满足。

Q11: The course was relevant and beneficial to my professional development.

该课程与我的职业发展相关，并对我的职业发展有益。

Q12: The course helped me look at life and the world with a new perspective.

该课程帮助我以新的视角来看待生活和世界。

Q13: The Instructor's knowledge was relevant and up to date.

老师的知识是相关的、最新的。

Q14: The Instructor came prepared for classes.

老师是有备而来的。

Q15: The Instructor taught in accordance with the announced lesson plan.

老师是按照公布的课程计划来授课的。

Q16: The Instructor was committed to the course and was understandable.

老师对课程很投入，而且很容易理解。

Q17: The Instructor arrived on time for classes.

教官准时到场上课。

Q18: The Instructor has a smooth and easy to follow delivery/speech.

指导老师的授课/讲话流畅、易懂。

Q19: The Instructor made effective use of class hours.

指导老师有效地利用了课堂时间。

Q20: The Instructor explained the course and was eager to be helpful to students.

指导老师对课程进行了解释，并渴望对学生有所帮助。

Q21: The Instructor demonstrated a positive approach to students.

指导老师对学生表现出积极的态度。

Q22: The Instructor was open and respectful of the views of students about the course.

指导老师对学生对课程的意见持开放和尊重的态度。

Q23: The Instructor encouraged participation in the course.

指导老师鼓励学生参与课程。

Q24: The Instructor gave relevant homework assignments/projects, and helped/guided students.

老师布置了相关的家庭作业/项目，并帮助/指导学生。

Q25: The Instructor responded to questions about the course inside and outside of the course.

指导老师对课程内外的问题作出了回应。

Q26: The Instructor's evaluation system (midterm and final questions, projects, assignments, etc.)

effectively measured the course objectives.

老师的评估系统（期中和期末问题、项目、作业等）有效地衡量了课程目标。 衡量课程目标。

Q27: The Instructor provided solutions to exams and discussed them with students.

老师提供了考试的解决方案，并与学生讨论。

Q28: The Instructor treated all students in a right and objective manner.

老师以正确和客观的方式对待所有学生。

Q1-Q28 are all Likert-type, meaning that the values are taken from {1,2,3,4,5}

Q1-Q28都是(Likert scale)**李克特量表類型的**，即从{1，2，3，4，5}中取值。

科普 Likert scale

是一種心理反應量表，常在問卷中使用，而且是目前調查研究（survey research）中使用最廣泛的量表。當受測者回答此類問卷的項目時，他們具體的指出自己對該項陳述的認同程度。

李克特選項是一個陳述。受測者被要求指出他或她們對該題目所陳述的認同程度，或任何形式的主觀或客觀評價。

通常使用五個回應等級，但許多計量心理學者（psychometrician）主張使用7或9個等級。一項最近的實證研究[3]指出，5等級、7等級和10等級選項的數據，在簡單的資料轉換後，其平均數、變異數、偏態和峰度都很相似。例如李克特的五等選項的：

1. 強烈反對
2. 不同意
3. 既不同意也不反對
4. 同意
5. 堅決同意

根據學生認定的難度我們也可以分為非常簡單 簡單 普通 困難 非常困難

**資料清洗**

資料清洗在資料處理中扮演著重要的角色。它包括檢測、修正和刪除資料集中的錯誤、不準確性、不完整性和不一致性。

資料清洗的重要性如下：

提高資料品質：資料清洗有助於提高資料品質，確保資料的準確性、完整性、一致性和可靠性。清除錯誤和不準確的資料可以減少對後續分析和建模的負面影響。

支持準確的分析和決策：清洗資料可確保所使用的資料合乎邏輯，消除了錯誤和不一致性。這有助於獲得準確的分析結果和做出可靠的決策。

處理缺失值：資料清洗涉及處理缺失值的問題。對於缺失值，可以採取填充、刪除或插值等方法，以便在後續分析中不會產生偏差或誤導。

上述主要提了一些對於我們此次專案有用到的重要性的部分做補充，

我們對此資料集總過做了3種清洗，缺失值處理 異常值處理 重複值處理

儘管先前認識資料集我們知道了他標榜沒有任何缺失值,

但是千萬不要相信任何人，於是我們還是需要自己檢查一次資料集

先檢查每一個欄位的資料是否均不為空

再來是要過濾掉所有數值都一樣的資料

科普：全同回答（Identical Responses）

為什麼要去除數值都一樣的資料？

在資料集中，某一特定變數（例如問卷中的問題）的回答在所有樣本中完全相同的情況。換句話說，所有受試者對該問題的回答都是一致的。

因為它們缺乏變異性和差異性，無法提供有關樣本之間的差異或特徵的有用資訊。這種情況可能是由於多種原因引起的，包括問題的設計、回答選項的限制、受試者間的合作或調查者的影響等。

全同回答可能表示資料品質的問題，例如受試者沒有真實回答問題、回答過程中存在錯誤或偏差等。在資料分析中，全同回答可能導致偏誤的結果或無法得出可靠的結論。

過濾完之後 結果能用的資料剩下一半

我們要先檢查我們的csv資料集中是否有超過範圍的資料

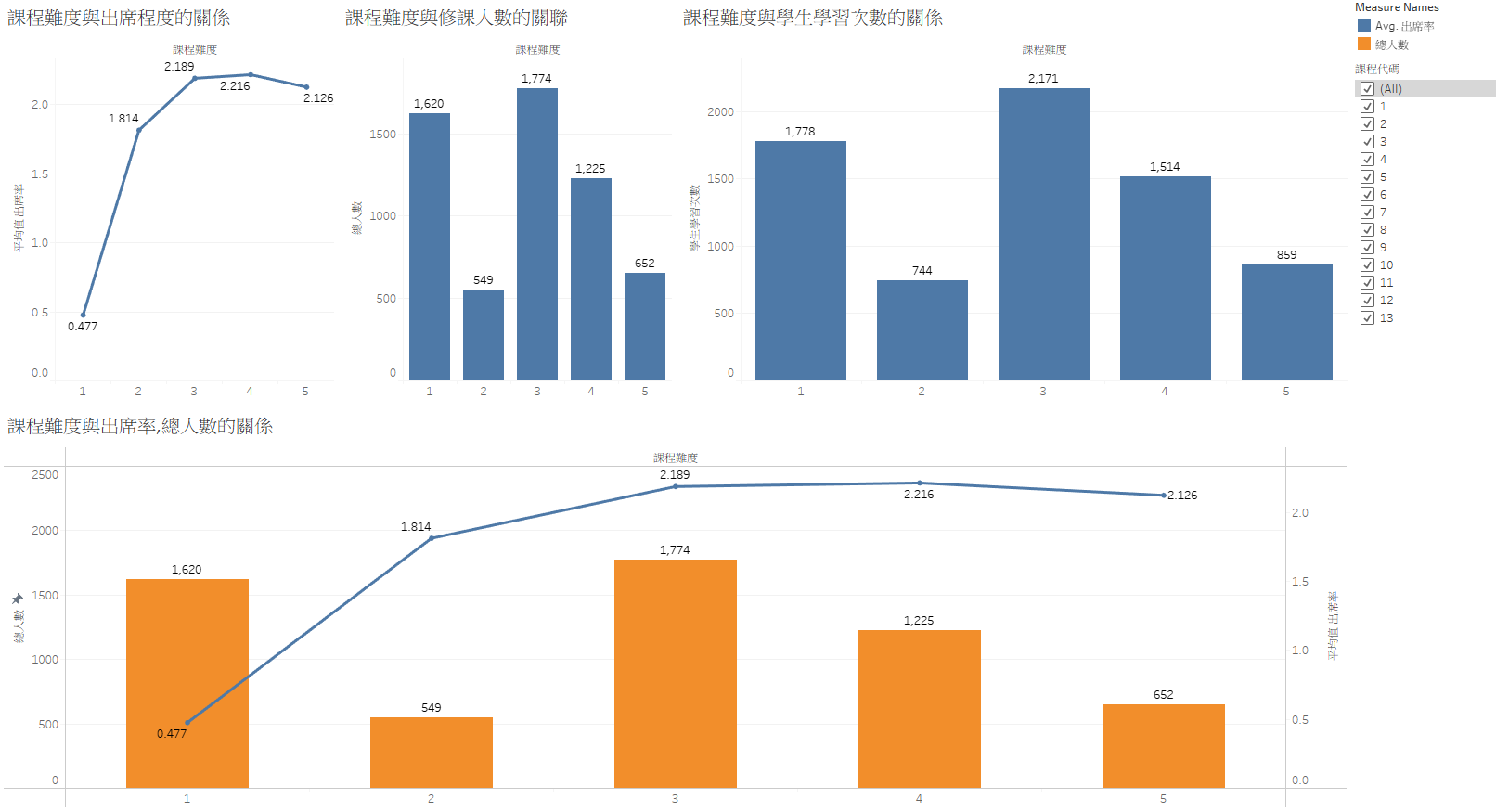
例如Q1 只能有1~5 但是他因為不知名的錯誤變成範圍外的數字，像是0或6，

不過尚未解決全同回答的資料仍有部分用途，比如說受試的總人數…等

資料清洗的Python程式碼

https://github.com/salasrew/NTUT\_DataScience/blob/main/datawashing.py

**資料視覺化**



（圖表一～三）

圖表一

課程難度與出席率的關係

可以看到當課程難度對於學生來說非常簡單的時候其出席的狀況最糟糕

隨著課程難度的增加，學生的出席率逐漸上升

圖表二

課程難度與學生學習次數的關係

出乎我們意料之外的部分是，課程難度非常簡單以及普通的學習次數竟為大宗。

我們認為可能的原因是因為該堂課要拿到滿分的機會特高，所以時間與成績的報酬率是相對高的

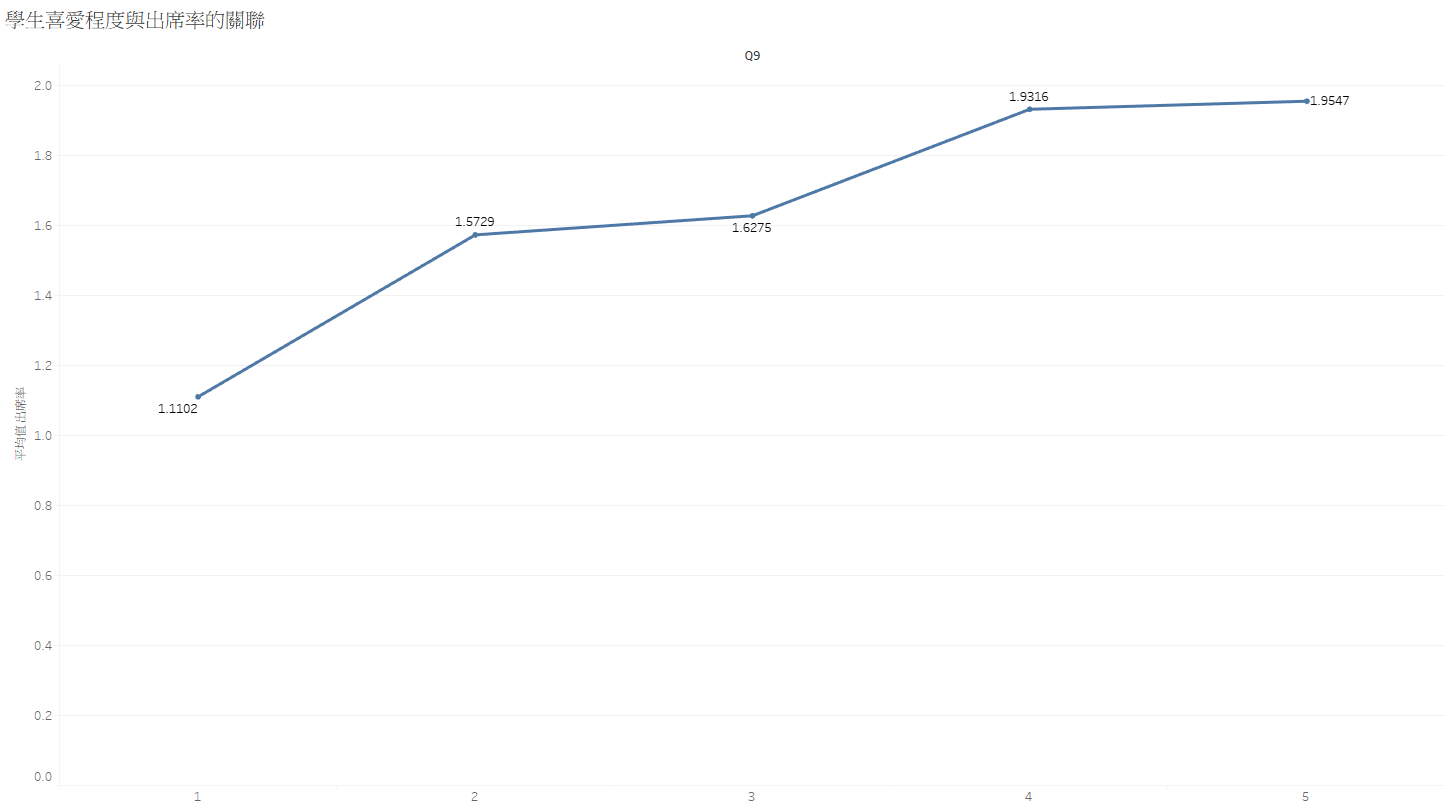
課程難度普通的時候，根據常態分佈來說符合大部分學生的能力，所以說難度適中讓學生有部分的壓力，人在適當的壓力之下能夠成長的飛越，激起學生對於學習的熱愛，所以說對學生難度普通的課程在學習上是很有幫助的。

圖表三

課程難度與修課人數的關聯

同圖表三的想法，我們也認為因為課程難度對於學生願意修該堂課的意願將大幅增加，

有學生會因為學分不足而無法畢業，這時候他們就會為了學分而去學習一些對自身沒什麼大幫助的課程。



（圖表四）

圖表四

根據Q9: I greatly enjoyed the class and was eager to actively participate during the lectures.

的資料，我們可以發現學生如果越喜歡這堂課，對於課程的參與率很直觀的越來越高

結論