作業3: 決策問題決定一切

**Q1. 典型的資料分析可以用以解決下面六種問題。**

Data analysts typically work with six problem types

1. Making predictions 2. Categorizing things 3. Spotting something unusual 4. Identifying themes 5. Discovering connections 6. Finding patterns

### **Making predictions**

Using data to make informed decisions about how things may be in the future.

A company that wants to know the best advertising method to bring in new customers. Analysts with data on location, type of media, and number of new customers acquired as a result of past ads can't guarantee future results, but they can help predict the best placement of advertising to reach the target audience.

### **Categorizing things**

Grouping data based on common features.

An example of a problem requiring analysts to categorize things is a company's goal to improve customer satisfaction. Analysts might classify customer service calls based on certain keywords or scores. This could help identify top-performing customer service representatives or help correlate certain actions taken with higher customer satisfaction scores.

### **Spotting something unusual**

Identifying data that is different from the norm.

A company that sells smart watches that help people monitor their health would be interested in designing their software to spot something unusual. Analysts who have analyzed aggregated health data can help product developers determine the right algorithms to spot and set off alarms when certain data doesn't trend normally.

### **Identifying themes**

Recognizing broader concepts and trends from categorized data.

User experience (UX) designers might rely on analysts to analyze user interaction data. Similar to problems that require analysts to categorize things, usability improvement projects might require help prioritize the right product features for improvement. Themes are most often used to help researchers explore certain aspects of data. In a user study, user beliefs, practices, and needs are examples of themes.

By now you might be wondering if there is a difference between categorizing things and identifying themes. The best way to think about it is: categorizing things involves assigning items to categories; identifying themes takes those categories a step further by grouping them into broader themes.

### **Discovering connections**

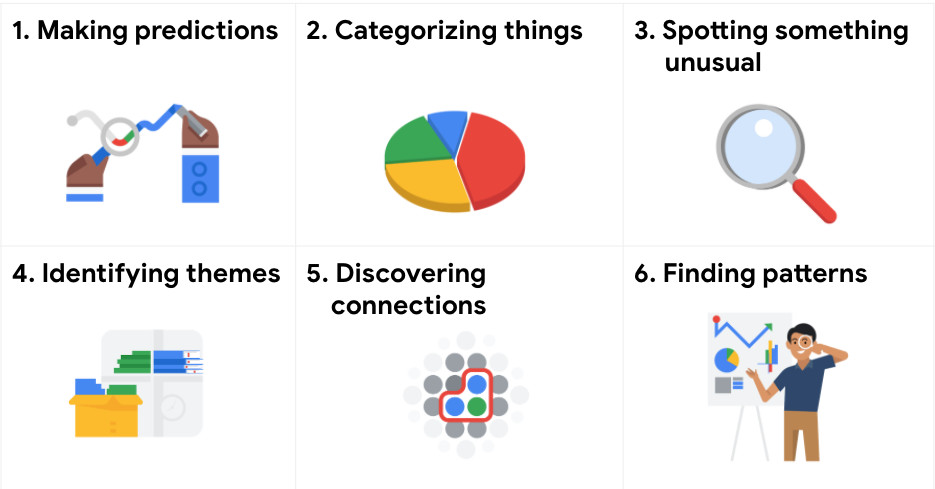
Identifying similar challenges across different entities—and using data and insights to find common solutions.

A third-party logistics company working with another company to get shipments delivered to customers on time is a problem requiring analysts to discover connections. By analyzing the wait times at shipping hubs, analysts can determine the appropriate schedule changes to increase the number of on-time deliveries.

### **Finding patterns**

Using historical data about what happened in the past to understand how likely it is to happen again.

Minimizing downtime caused by machine failure is an example of a problem requiring analysts to find patterns in data. For example, by analyzing maintenance data, they might discover that most failures happen if regular maintenance is delayed by more than a 15-day window.



接下來，請依照您抽到的牌卡順序，各給出例子，並應用可能的open data資料說明其輸入和輸出資料。

例如:

1. Making predictions（行）：

* 輸入：[即時交通事故資料(A1類) | 政府資料開放平臺 (data.gov.tw)](https://data.gov.tw/dataset/12818)
* 輸出：預測交通事故風險，使用預測模型對未來可能發生事故的地點和時間進行預測。
* 決策：制定交通管制、監測和安全提示等策略，減少交通事故風險。

1. Categorizing things（老）：

* 輸入：[臺南市各區長照需求人口推估表 - 臺南市111年度各區長照需求人口推估表 - 臺南市政府資料開放平台 (tainan.gov.tw)](https://data.tainan.gov.tw/dataset/ltc_estimate/resource/17236752-a1ba-48fc-a8c7-3b1bd20b07ef)
* 輸出：分類區域是否具有高比例的65歲以上失能老人，透過將人口以 "65歲以上失能老人" 欄位分成兩類，假設500人以上為高比率，500人以下為低比率。
* 決策：如果地區65歲以上失能老人的比例較高，則相關機構可以考慮增加該地區的居家照護服務和社區照顧中心等措施，以支持這些老人在家中生活。

1. Spotting something unusual （病）：

* 輸入：[台灣COVID-19冠狀病毒檢測每日送驗數 | 政府資料開放平臺 (data.gov.tw)](https://data.gov.tw/dataset/120451)
* 輸出：使用時間序列分析來檢測病毒數量隨著時間的變化是否有異常高峰或下降。同時，比較陽性測試結果和總測試量，以判斷陽性率是否有異常的變化，這可能意味著病毒的傳播率發生了變化。
* 決策：及早發現疫情對於採取快速有效的公共衛生措施阻止病毒傳播非常重要。這能讓公共衛生官員盡快採取隔離、接觸者追踪、檢疫措施和增加檢測等干預措施，以打破傳播鏈，防止進一步傳播。

1. Identifying themes （衣）：
2. Discovering connections （育）

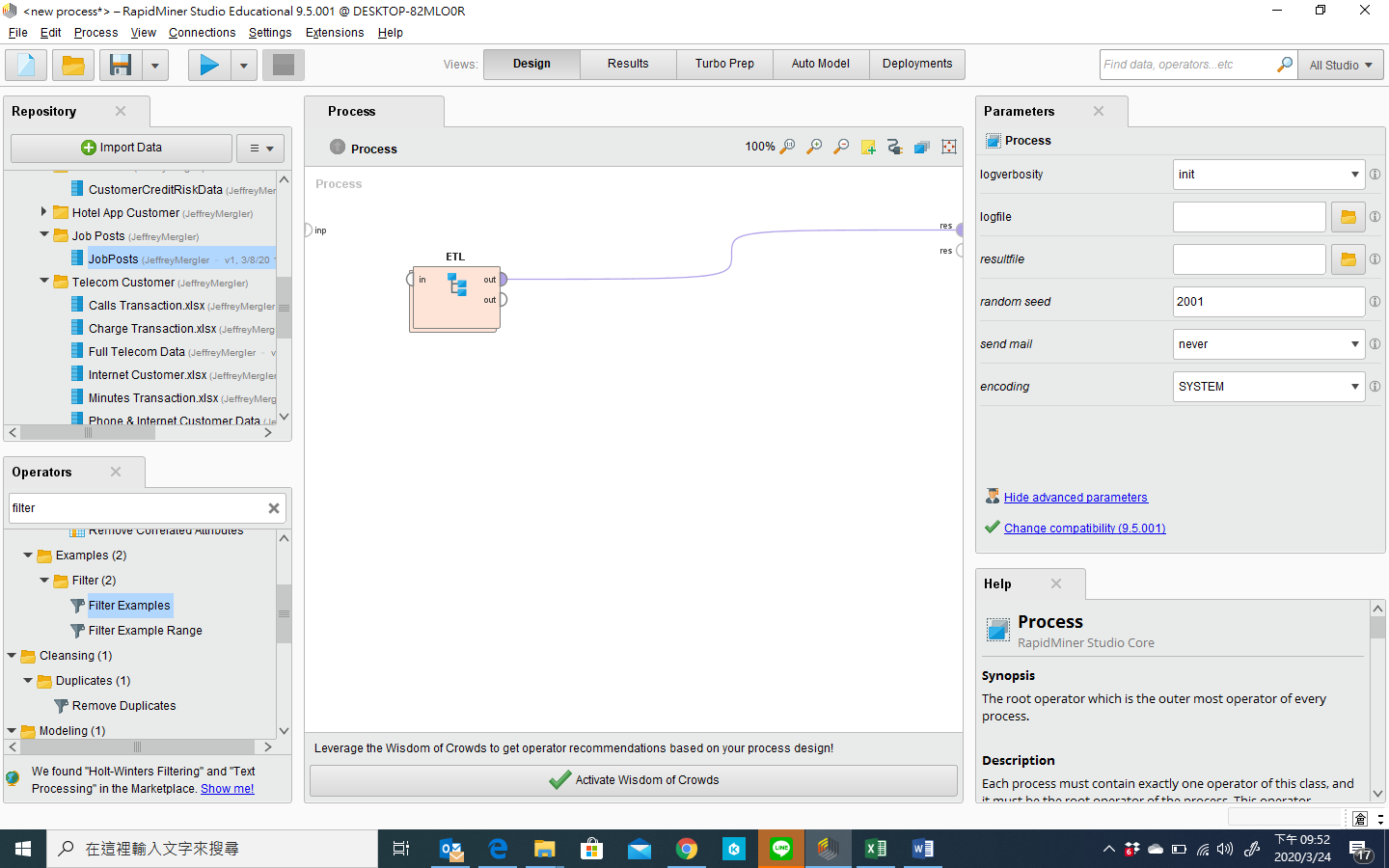
* 輸入：各級學校分布位置、[臺北市各學校可上網載具設備配備情形](https://data.gov.tw/dataset/135815)
* 輸出：網路設備配備情形和學校的分布位置之間的相關性

1. Finding patterns（樂）：

* 輸入：[Daily Box Office For 2023 - Box Office Mojo](https://www.boxofficemojo.com/date/?ref_=bo_nb_shs_tab)
* 輸出：預測電影季節性趨勢，分析資料以確定電影收入是否存在任何季節性趨勢。
* 決策：如果分析結果顯示某些月份或季節電影收入通常更高，投資者可能會考慮投資於這些時間上映的電影。此信息還可以幫助投資者有效地分配預算並優化投資組合，以利用季節性趨勢。如果分析顯示某些季節或月份對電影業來說通常利潤較低，投資者可以調整他們的投資策略，以減少在這些時期的風險。

**Q2 薪資大不同**

2-1 讀入檔案(salary.csv),請大概說明一下您對於資料的了解(5%),並進行「必要」之資料前處理,並請將相關前處理存於ETL子流程如下圖(請說明您做了什麼,為什麼) (10%)



2-2 .請試著解決並以salary為預測標的，貼出您的決策樹,並加以說它的意涵(10%)

2-3 請列出您的混肴矩陣,並請加以說明其意涵，除此之外，我還希望您能列出您模型的下述４個指標，並請簡略說明他們的計算方式及含意(5%)



2-4最後，請說明您的管理意涵？ (10%)

請將您的process存檔為學號-1, 如: 106AB001\_1.rmp檔