**題目: COVID-19在台分析**

**對零售業和餐飲業的影響**

**組員:**

**巨資二A 07170121許哲聞**

**巨資二A 07170133 彭鈺軒**

**巨資二A 07170138 簡呈澔**

**會三B 06152243 曹詠喻**

**社四B 05115246 黃聖崴**

**目標:**

我們的報告主要探討Covid-19對臺灣人民的影響，包括: 「各年齡的確診情況」、「各縣市的確診情況」以及「各縣市的確診人數與各縣市人口密度的相關性」，並試圖提出針對疫情大家應該要注意的是什麼以及如何避免感染。此外疫情除了衝擊人之外，對於民生經濟的摧殘也是時有耳聞，因此我們也會進行

「SARS與COVID-19疫情期間的民生經濟趨勢分析」來比較不同時期疫情的衝擊狀況。經過討論後，我們打算以餐飲業與零售業的營業額作為民生經濟的代表行業。

**背景介紹:**

武漢肺炎爆發，世界各國開始陷入警戒狀態。各領域開始積極地研究疫情的狀態。台灣的防疫良好受到世界關注，我們不禁也想，如何透過自身所學來協助台灣防範疫情，此外也想了解並分析疫情對臺灣民生的影響，經過討論後，我們決定對零售業和餐飲業進行影響評估。

在2003年台灣有類似的疫情爆發經驗，而那年的疾病為SARS。因此，我們想了解SARS與COVID-19對零售業和餐飲業的影響強度有多深，進而推測COVID-19的恢復期要多久。Covid-19的流行期間為2020年1月~至今，我們手中的資料只到4月，因此，我們採取的時間為1月~4月，我們將以零售業和餐飲業的營業額作為為依據來了解這兩段時間的影響。

**資料來源:**

1. 政府資料開放平台: 批發、零售及餐飲業營業(本資料紀錄的期間為民國88年至民國109年4月)
2. 衛服部疾管署open data: SAS與COVID-19的資料
3. 台灣各縣市人口密度
4. 國人海外工作普查

**分析計畫:**

資料整理:

1. 不合格式的值、缺失值修正

針對資料的缺失值和格式不符合我們需求的資料進行調整。首先針對SARS和COVID\_19的資料其年齡(age)欄位在進行由小至大排序時，個位數年齡的區間無法正確地被排在最前面，因此在數字前添加0。例如5-9歲會變成05-09歲，以此讓資料能被正確的排序，此外這兩份資料皆無缺失值。

此外零售與餐飲業的資料，我們將營業額的單位從千元轉換為億元，這樣比較能一目了然的判斷。

台灣各縣市人口密度以及國人海外工作普查等資料並無缺失或不符合我們期望的值。

1. 篩選資料

為了針對COVID\_19現況了解，我會根據年齡層(age)、縣市(city)進行確診人數的加總。

海外工作人口普查是為了配合境外移入的確診案例僅參考總人數來分析。台灣各縣市人口密度COVID\_19零確診的城市會被篩掉。

而關於零售和餐飲業的資料，我們從批發、零售及餐飲業營業的資料中，將零售和餐飲業篩選出並分成不同的資料表。並取出兩種行業疫情期間以及其前一到二年同時期的資料，SARS期間我們僅篩選了2001~2003年3月到7月的資料。COVID\_19則是2018~2020年1月到4月的資料來進行分析。

變項整理:

1. COVID\_19現況分析

加總確診案例數、根據年齡層、性別、是否為境外移入及各縣市計算確診人數、以及確診人數的比例。

1. 零售業、餐飲業分析

為了使營業額單位變得更清楚，我們將營業額的單位千元轉換成億元。並根據不同月份進行營業額的加總，以此來繪製折線圖並分析。

**結果與討論:**

COVID\_19現況分析:

根據這圖(一)我們可以很清楚的看到各年齡層的確診人數占比(age\_fp)，可以很明確地發現20~34歲的確診人數占比為最大宗，這顛覆了「老年人和幼童免疫力差容易染病」的想法。

由數據可知，儘管在醫學上，年紀太小或高齡皆是免疫力較差的族群，然而感染風險還是要看接觸環境，也就是在戶外的活動時長。因此20~34歲的青壯年族群成為最主要的發病族群，可得知減

少戶外活動能夠降低被感染的風險。



圖(一)



根據圖(二)我們可以發現直轄市的確診人數占比(city\_fp)都是偏大的，這使我們決定找出原因。而我們發線這些縣市的共通點都是大量人口聚集地，因此我們決定要針對人口密度與確診的關係進行調查。

圖(二)



圖(四)

圖(三)

根據圖(三)的排序我們認為人口密度(pop\_d2)與疫情的相關性是存在的，為了更進一步驗證，我們進行了人口密度與各縣市確診人數(city\_f)的相關分析，最後也驗證我們的推測結果如圖(四)，p值<0.05達顯著水準，可得知人口密度與確診人數有相關，並且強度為0.63的中偏強的相關。由此可知，人口密度高的縣市要多避免去人口聚集處，如觀光景點類的，才能避免感染、降低風險。

圖(五)是根據境外移入與性別為根據區分的確診占比(overseas\_sex\_fp)首先可以看出境外移入佔了確診人數的大宗，此外在男女的確診占比上也都接近1:1，可知COIID\_19不會偏好感染特定性別。

由此也可發現台灣的確診案例幾乎都是境外移入，可見台灣本土的防疫在許多人的努力下非常有成效。

圖(五)

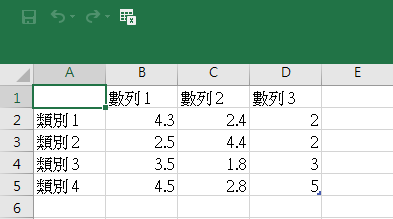
 圖(六)為死亡的案例數，雖然不多但可以發現死亡的都是中高齡人口，也就是說抵抗力較差的中高齡族群需要多注意。

圖(六)

疫情對民生經濟衝擊分析:

我們以疫情爆發的前兩年作為比較因此SARS的部份我們將以90年到92年3月到7月的營業額以折線圖來呈現。而COVID\_19則是107年到109年1月到4月。

由於零售與餐飲業的資料，其日期格式為民國年，我們利用SAS在轉換民國至西元年上遇到技術困難導致繪製出的折線圖都有問題，若將各年分開畫不能達到我們想要的視覺化效果，不利於分析，因此我們選擇將表單資料分別匯出後，利用PPT內建的功能來完成折線圖。首先點選插入->圖表->折線圖。右鍵預設的折線圖，再點選編輯資料，就會出現EXCEL表單，便可以開始設定我們的折線圖資料。如下圖(七)。



圖(七)

在欄位A，填入月份，例如: SARS的疫情期間是3月到7月；Covid-19的疫情期間是1月到4月，就可以將這兩段期間分別填入欄位A。B1到D1填入年份，例如: SARS的盛行年分是民國92年；Covid-19的盛行年分是民國109年，我們組取疫情發生的前兩年進行比較，並且將這兩段期間分別填入B1到D1。最後將零售業和餐飲業的營業額分別填入，將EXCEL表單填完後，PPT就會自動呈現出折線圖，我們以此種方式繪製出折線圖並分析。

圖(八)

從圖(八)的趨勢可看出在3月到6月的期間民國90年到民國91年的曲線呈現的非常相似，除了民國91年7月的成長率不如民國90年，民國91年的月平均成長率仍為2.82%，雖然漲幅不大，但我們可以根據這兩年的資料來推斷下一年的營業額是會呈現正成長。到了民國92年(SARS爆發的當年)，民國92年的折線與民國91年相比整條向下移動。6月和7月的確診人數分別為14人和0人，我們組認為那兩個月有突破性的成長是因為疫情趨緩，而產生的「消費意願回升」，盡管如此，民國91年和民國92年的月平均成長率仍為-7.49%，可見疫情對餐飲業的影響。

為了更精準的證明我們的分析，我們對SARS期間的餐飲業營業額與確診人數進行相關分析如下圖(九):

圖(九)

由上可知其p值<0.05為顯著相關，關係為強度的副相關。也就是說疫情導致了餐飲業營收的負向成長的關係。

圖(十)

由圖(十)可以看出，在SARS發生前零售業呈現穩定的成長（平均成長率4.92％），且有一定的季節性，3~4月由於剛過完年消費力趨緩，到了四月慢慢地開始回升，當報稅季結束也就是6月後，消費力明顯攀升，而我們對比SARS當年的數據發現，當疫情開始時，並沒有影響零售業該有的銷售額，仍穩定成長（對比去年），但原本應是回升的4月卻明顯往下走，甚至低於上一年的水平，這很可能是因為當時的疫情來到了高峰，大家開始大幅減少出門的次數，連帶影響了整個零售業，不過當疫情趨緩（6月），零售業很快的就回到了該有的水準，以整體而言影響並不是特別大（平均成長率2.53％）。

圖(十一)

從圖(十一)我們可以看出餐飲業的趨勢，107到108年的走勢都大概是1到2月為成長期，2月過後開始下跌，然而疫情爆發之後可以看到從1月開始到4月都呈現快速下跌的狀態，可見疫情對於餐飲業的衝擊在趨勢上是明顯的。為驗證我們的想法，我們也進行了營業額與確診人數的相關分析如下圖(十二):



圖(十二)

可發現在COVID\_19期間營業額與確診人數的相關性變弱了，不僅不顯著，連強度也削弱了，我們認為是因為109年開始外送平台如uber和foodpanda等崛起，使得營業額受到的影響不如SARS期間。

圖(十三)

從圖(十三)我們可以看到這三年的走勢圖是差不多的，雖然我們在技術上限制沒辦法做更精準的趨勢分析，但通常每一年的經濟成長趨勢都有一定的模式可循，以視覺化的折線圖來看，其形狀會差不多，除非有重大事件影響例如我們所分析的疫情引響，然而從圖上看來其實疫情的影響可說是非常微小，為了證明我們的推測，我們針對零售業的109年的營業額與確診人數進行相關分析，得出圖(十四)的結果如下:



圖(十四)

可以看出確診人數對營業額並無顯著相關。也就是疫情衝擊不太會影響零售業的營業額。

**結論:**

根據上一章節，結果與討論中我們整理出了以下問題，並一一回答作為結論。

1. 年紀大或年紀太小是被感染的高風險族群?

由數據可知，儘管在醫學上，年紀太小或高齡皆是免疫力較差的族群，然而感染風險還是要看接觸環境，也就是在戶外的活動時長。因此20~34歲的青壯年族群成為最主要的發病族群，可得知減少戶外活動、保持社交距離能夠降低被感染的風險。然而從死亡率來看的確高齡族群較高。

1. 什麼情況下感染的機率最高?

從人口密度與確診人數的相關分析可得知。在人潮擁擠的地方感染機率最高，無論是數據或是醫學上都得證，因此建議暫時遠離人群，以網路代替。

1. SARS和COVID\_19哪個對民生經濟衝擊比較嚴重?

以零售業而言，兩者影響力並無顯著差距，但透過圖表，我們可推論，疫情在最嚴重的時期，零售業營業額依然會受影響。

以餐飲業而言，SARS的影響較為嚴重，透過圖表，我們發現武漢肺炎的折線斜率相較SARS平緩。我們推論在2002年SARS爆發時，因為尚無食物外送平台，所以SARS對餐飲業造成的影響較顯著。

1. 以SARS趨勢圖來看COVID\_19的恢復期大概要多久?

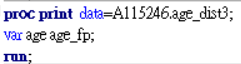
根據SARS疫情後恢復情況，我們推測一旦疫情趨緩，經濟就會開始復甦，但月營收成長率圖表顯示，COVID-19影響時間較長，因此恢復正常的時間會比SARS長。我們預估被武漢肺炎的影響經濟將於疫情趨緩的6月開始復甦。

1. 振興券是否該針對特定產業發放?

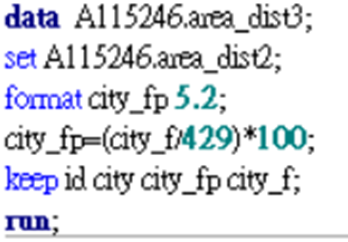
藉由比較零售業和餐飲業受到疫情的影響，可得知餐飲業受到的影響更為顯著，因此，我們認為相較於零售業，餐飲業更需要振興券的幫助。當然只要遇到疫情最慘的就是旅遊業，但若是已經知道疫情衝擊他們的營業額一定會跌入谷底，可能比較沒有分析的價值因此我們沒有選擇旅遊業作為分析對象，況且旅遊比較偏娛樂、觀光業，非民生必要。

**附錄:**

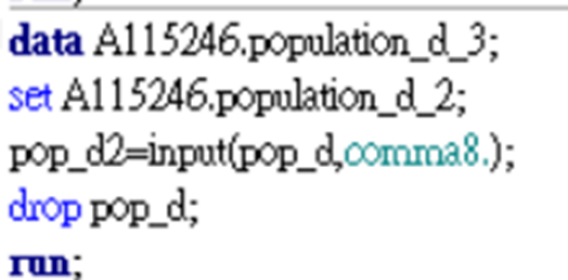
關於程式碼:



在整理完資料後創一個新的資料age\_dist3，並取出其中兩個欄位來得知各個年齡層的確診人數，並將結果製成圖(一)。

****

建立資料 A115246.area\_dist3來了解各個縣市確診人數佔的比例，並將結果製成圖(二)。



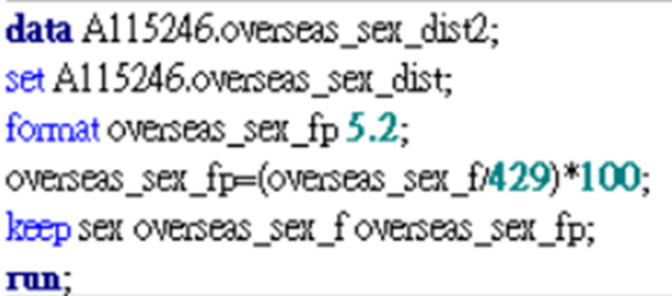
經由轉換格式後，再整理A115246.population\_d\_3得知各縣市的人口密度，並將結果製成圖(三) 。



我們先用left join合併A115246.area\_dist和

A115246.population\_d\_3之後，再用proc corr計算相關係數，並將結果製成圖(四) 。

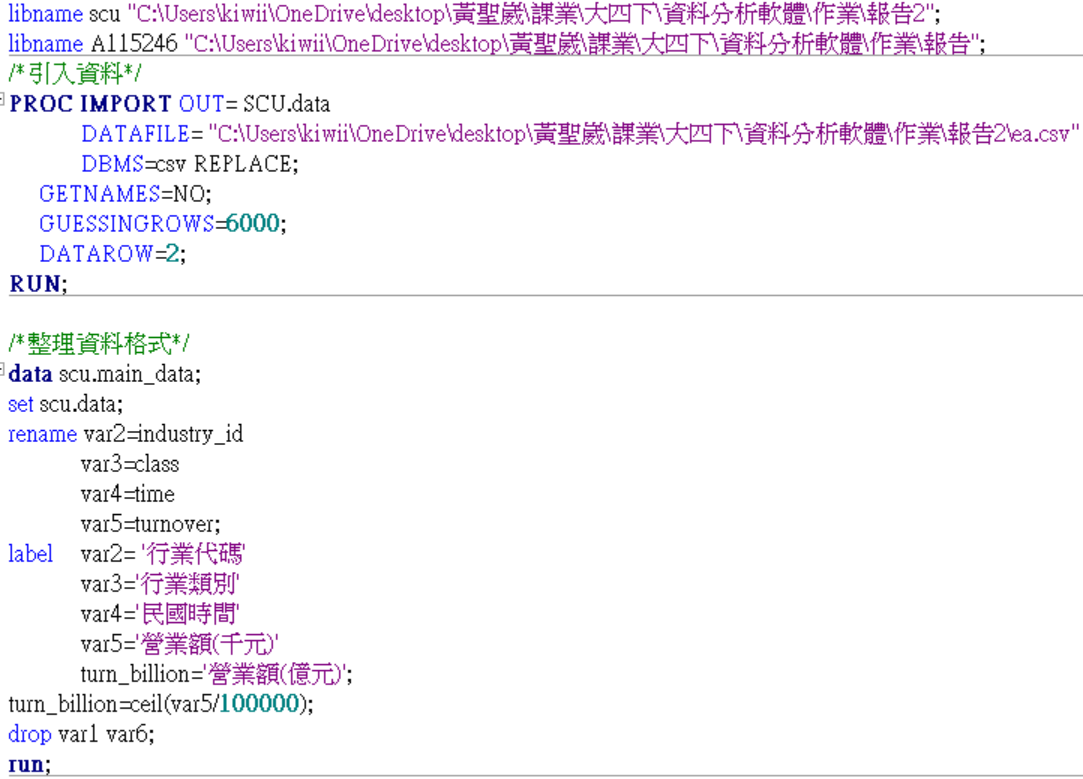




我們建立A115246.age\_overseas\_dist來看境外的狀況，並將結果製成圖(五) 。

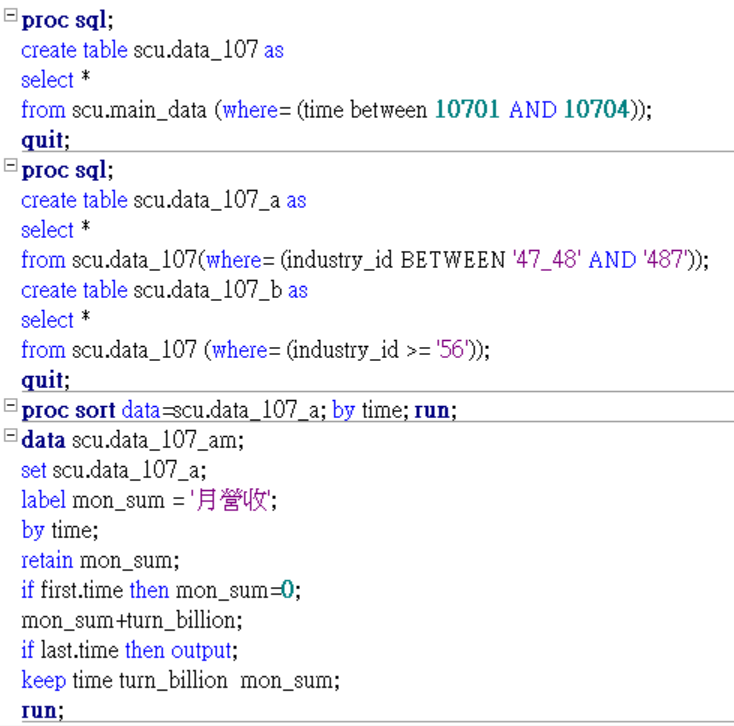
最後透過proc freq將A115246.death1做累計次數分析，並做成圖(六) 。

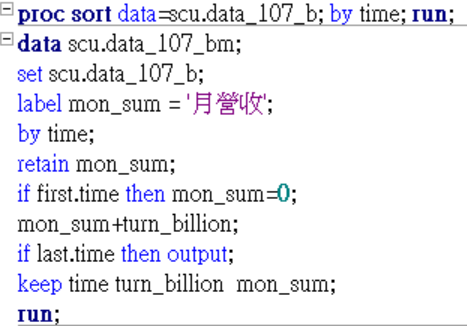




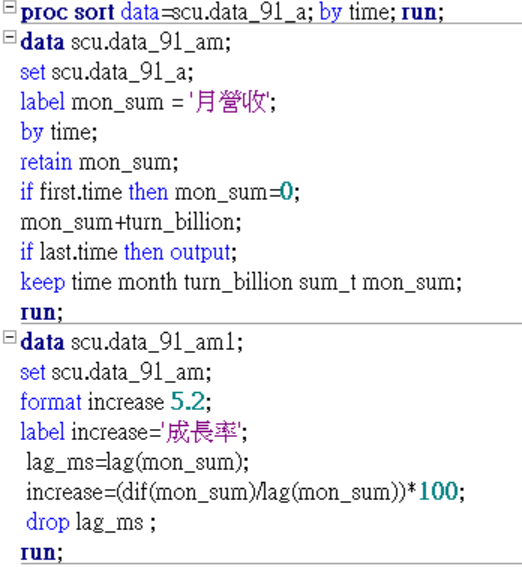




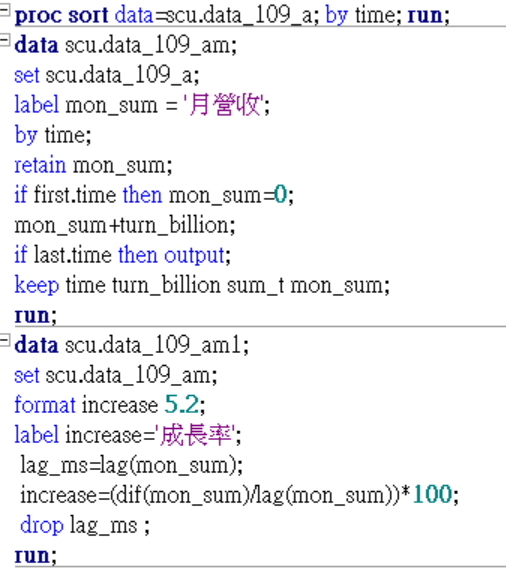






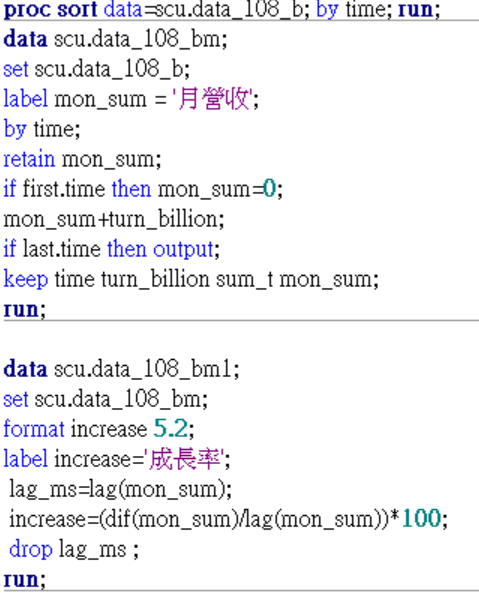


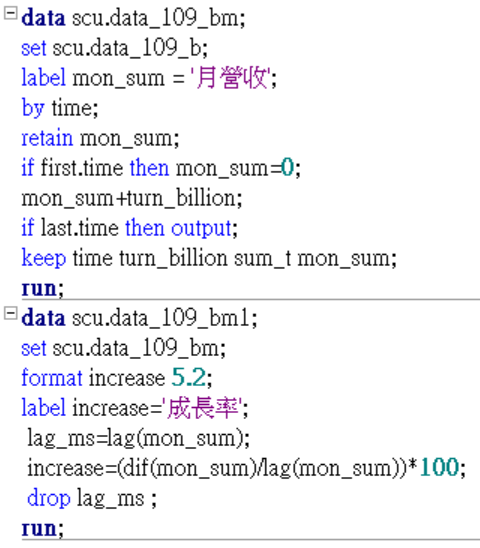


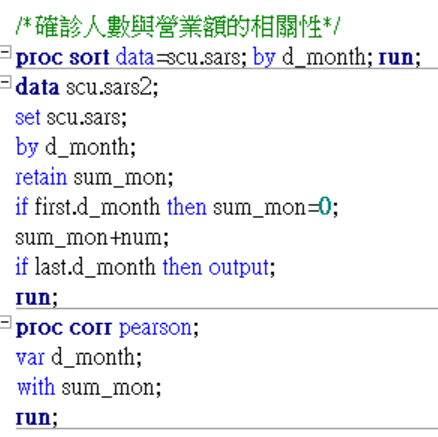


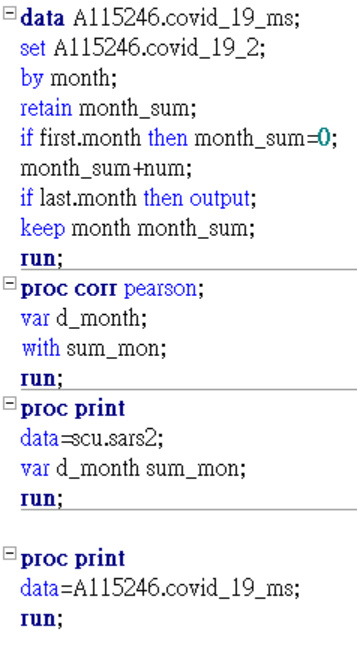












執行過程發生的問題:

(1) 武漢肺炎還在發生中，因此資料有限。

(2) 處理圖表時民國年的資料再轉換成西元的過程有困難。

(3) 由於民國轉西元的問題，我們無法用SAS畫出不同年疊層的折線圖改使用

excel輔助。

(4) 匯入資料出現亂碼。

延伸的問題與展望:

1. 針對兩岸以及各國對中國的態度，因為這次疫情有非常大的變化，針對社

群研究也許可以朝這個方向蒐集資料。

1. 疫情對於餐飲業的衝擊也隨著網路及新興產業的出現(如外送業)而減小。
2. 不論是武漢肺炎或是SARS，都確實對餐飲業與零售業造成影響。經濟是鏈

狀結構，一旦牽一髮則動全身，使得受到衝擊的不僅僅是餐飲與零售業，而是對整體經濟造成影響。