# 

# 2004-NPcrestHori_B-W

# School of InfoComm Technology

**Data Exploration & Analysis Assignment**

Diploma in DS

April 2023 Semester

**ASSIGNMENT 1**

(30% of DEA Module)

7th May 2023 – 4th June 2023

**Submission Deadline:**

**Presentation: 4th Jun 2023 (Sunday), 11:59PM**

**Excel files: 4th Jun 2023 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **T02** |
| **Student Name** | **:** | Fionn Lim |
| **Student Number** | **:** | S102400073K |

**Penalty for late submission:**

10% of the marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 11th Jun 2023 (Sunday), 11:59PM.

# 1. OBJECTIVE

To assess students’ ability to research and work on semi-structured and unstructured tasks and problems by applying their knowledge of Microsoft Excel functions and feature exploration in an integrative approach.

# 2. SCOPE

This is an individual assignment. The assignment consists of 3 practical questions. These questions are to expose students to other possibilities of using spreadsheets to address problems in other businesses or industries, to work on personal productivity projects, and exploring a financial dataset. You will be applying the knowledge of Microsoft Excel functions and feature exploration covered in Week 1 to 6.

**Your answers to all questions must be formatted suitably and accompanied with relevant documentation/explanatory notes where appropriate.**

**IMPORTANT: Plagiarism and Copyright Issues**

Plagiarism means, “copying any part of a source, and then submitting it, claiming that it is your own work.”

Please ensure that all the works submitted by you are not copied from other sources. Any attempt to plagiarize will be dealt with severely, and it may result in your failing the module.

If you have made any references to certain materials, make sure you **cite the sources** by acknowledging and providing the information necessary to find the source (e.g. Title and author of book, Internet links, etc.)

Please refer to the following URL for more details:

<https://www1.np.edu.sg/clte/antiplagiarism/policy.htm>

# 3. PRACTICAL QUESTIONS

## Question 1: Hotel Demand Booking (40 marks)

You are a customer relationship management analyst for a hotel analysing the hotel booking demand dataset for the year 2022, extracting insights on guest trends and behaviour.

The survey data has been collated on **HotelSurvey\_[student\_name]\_[student\_ID].xlsx** in the worksheet Survey Data. The country code used in Survey Data is given in the worksheet Country Code. The features are listed below:

|  |  |
| --- | --- |
| **Hotel** | Resort Hotel or City Hotel |
| **Lead Time** | Number of days that elapsed between the entering date of the booking into the hotel management system and the arrival date |
| **Arrival Date** | Date of arrival of guest |
| **Duration of Stay** | Number of Nights that the guest stays |
| **Country** | Country code on nationality of guest |
| **Number of Adults** | Number of adults in reservation |
| **Number of Children** | Number of children in reservation |
| **Number of Babies** | Number of babies in reservation |
| **Meal** | SC (Self Catering) – No meals are included; however, your accommodation will be provided with catering facilities for you to cook light meals.  BB (Bed and Breakfast) – Breakfast is included  HB (Half Board) – Breakfast and evening meals are included. In some cases, you can choose to receive lunch instead of breakfast – the hotel will confirm this on arrival.  FB (Full Board) – Breakfast, lunch and evening meals are included. |
| **Repeated Customer** | 0: No  1: Yes |
| **Days in Waiting List** | Number of days in waiting list |
| **Required Parking Space** | Number of car park lots requested |
| **Cancelled Booking** | 0: Booking not cancelled  1: Booking cancelled |

For each sub-question below, you are to create a new PivotTable report on a **fresh worksheet** (rename your fresh worksheet based on the sub-question e.g. **Q1a, Q1b, Q1c, Q1d**):

1. Using suitable **table and visualization charts,** contrast between the percentage of returning guest[[1]](#footnote-1) and the percentage of cancelled reservations[[2]](#footnote-2) among City hotel and Resort Hotel in the year 2022.

Average function

1. Based on the data from Resort Hotel, present (using **table and visualization)** the **top 5 months** with the largest expected guest crowd[[3]](#footnote-3) together with a breakdown of the crowd (adult, infant and children).
2. Provide **descriptive statistics** of the lead time (the time between booking and arrival date) for those arrival dates in the month of April 2022. Explain your interpretation on the statistics of lead time.
3. Using suitable **pivot chart and a visualization,** provide ananalysis on the crowd that you will expect from the various countries in the month of April 2022. Note: consider only those who did not cancel their bookings during that period.

## Question 2: Air Passenger Survey Insights (40 marks)

You are a customer analytics executive in the marketing department of an international airline. The head of the marketing department has recently engaged a third-party market research firm to conduct a survey on a random sample of 100 air passengers (comprising of 50 male and 50 female passengers) who have travelled with the airline for the past month. The survey data has been collated on **Air Passenger Survey\_[student\_name]\_[student\_ID].xlsx** in the worksheet **Survey Data**, and sent to you for further analysis. Your head of department has instructed you to create suitable PivotTable reports which can help answer some questions in relation to current customer trends and behaviour.

For each sub-question below, you are to create a new PivotTable report on a **fresh worksheet** (rename your fresh worksheet based on the sub-question e.g. **Q2a, Q2b, Q2c**):

1. Show the percentage breakdown and average age for each in-flight activity among all female passengers
2. For Economy class passengers, show the percentage breakdown and average annual income of the following age groups:

* 15-24
* 25-34
* 35-44
* 45-54
* 55-64
* 65-74
* 75-84

1. For Business class passengers, show the number of passengers and the average annual income for each preferred booking channel, split by nationality. In addition, use a slicer to filter only for those which indicated Internet or Mobile as their preferred check-in method.

## Question 3: Data Exploration on Credit Card Churn Dataset (20 marks)

## Credit card churning is the process of opening cards for the sole purpose of earning welcome bonuses or other benefits. Usually, it involves closing cards after the bonus posts to your account and before the next annual fee is charged[[4]](#footnote-4).

## You are a manager at the bank and are disturbed with more customers leaving their credit card services. The data set Creditcardchurn\_[student\_name]\_[student\_ID].xlsx collated a record of the existing card customer of your bank and those who have churned.

## You will now use your data exploration skills to study the characteristics/traits of customers who had churned, so that you can identify your existing customers who may potentially churn.

The features are listed as follows:

|  |  |
| --- | --- |
| **Feature name** | **Description** |
| **Customer\_Age** | Customer's Age in Years |
| **Gender** | M=Male, F=Female |
| **Dependent\_count** | Number of people dependents |
| **Education\_Level** | Educational Qualification of the account holder |
| **Marital\_Status** | Married, Single, Unknown |
| **Income\_Category** | Annual Income Category of the account holder |
| **Card\_Category** | Type of Card (Blue, Silver, Gold, Platinum) |
| **Months\_on\_book** | Time (Months) of Relationship |
| **Total\_Relationship\_Count** | Total no. of products held by the customer |
| **Months\_Inactive\_12\_mon** | No. of months inactive in the last 12 months |
| **Credit\_Limit** | Credit Limit on the Credit Card |
| **Total\_Trans\_Amt Num** | Total Transaction Amount (Last 12 months) |
| **Total\_Trans\_Ct Num** | Total Transaction Count (Last 12 months) |
| **Avg\_Utilization\_Ratio** | Average Card Utilization Ratio |
| **Churn** | 1: Customer churned  0: Existing customer |

You are to perform data exploration on the given dataset using the techniques discussed over the weeks. You are to explore the data in excel or other resources and discuss the data in univariate, bivariate and multivariate perspective. Write a report with supporting visualization on the findings during data exploration. Explain in detail the choice of visualization and critical findings. Include all resources used including websites and tools. You are to use the cover page of this document as the cover page of your report. Please refer to the table below as a guideline to craft your report.

|  |  |  |
| --- | --- | --- |
|  | **Suggested Report Sections & Content Guidelines** | **Word Count** |
| 1. | Table of Contents | NA |
| 2. | Abstract/Overview | 500 words |
| 3. | * Problem Understanding * Data Exploration * Univariate analysis * Bivariate analysis * Multivariate analysis * Others | Min: 1000 words  Max: 3000 words |
| 4. | Summary   * Summarize your findings * Reflections on Data Exploration and Analysis process | Min: 500 words  Max: 1000 words |

# 4. ASSIGNMENT DELIVERABLES

**Files submission**

Each student is required to submit all the required solution files in the assignment submission folder in BrightSpace. Before submission to BrightSpace, please ensure that the workbook solution files follow the naming convention below with the name and ID of the student.

The required solution files for submission are:

Q1: **HotelSurvey\_[student\_name]\_[student\_ID].xlsx**

Q2: **Air Passenger Survey\_[student\_name]\_[student\_ID].xlsx**

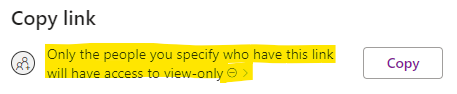
Q3: **Creditcardchurn\_[student\_name]\_[student\_ID].xlsx** and **Q3\_report\_[student\_name]\_[student\_ID].docx**

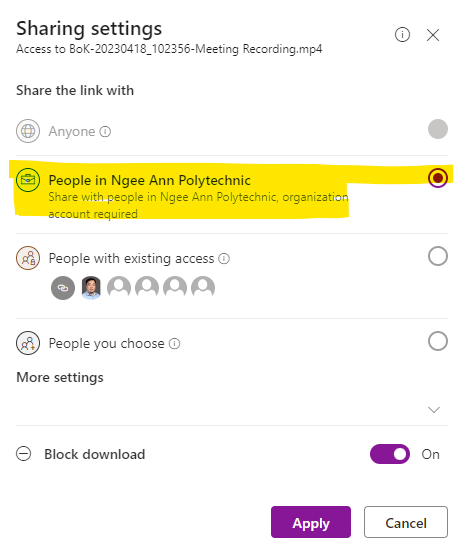
Zip all the files in a single folder and name the zipped folder as **[student\_name]\_[student\_ID].zip.** Submit the zip file to BrightSpace in Assignment 1 Submission by **4 Jun 2023 (Sun), 11:59pm**

**Video Presentation**

For the presentation, each student will prepare a video recording of no more than 10 mins of duration in MS Teams, to explain how the solutions were arrived.

The **link to your recorded presentation**

* Submit the link of your recorded presentation in Assignment 1 Submission. You can get the link by clicking  on the top right of the video opened in the web browser. Click on and click on the highlighted text to choose the sharing settings to **People in Ngee Ann** Polytechnic and click **Apply** as shown below. Copy the link and insert it in the **Comments** section of your submission.



* Extend your Teams recording expiration date to **No expiration**
  + - Open your video recording in a browser and at the bottom left of the screen to choose **No expiration**.
* **Note**: if tutors cannot access the recorded video or setting the expiration date to No expiration, and students need to be reminded by tutors to reset permission, students will be penalized for non-conformance.

**End of Document**

1. Defined as the (Number of repeated customer)/(Total number of reservations in 2022)\*100% [↑](#footnote-ref-1)
2. Defined as the (Number of cancelled reservations)/(Total number of reservations in 2022)\*100% [↑](#footnote-ref-2)
3. Expected Guest Crowd is the total number of guest (who has not canceled their reservation) arriving in the hotel. [↑](#footnote-ref-3)
4. <https://www.forbes.com/advisor/credit-cards/what-is-credit-card-churning/#:~:text=Credit%20card%20churning%20is%20the%20process%20of%20opening,and%20before%20the%20next%20annual%20fee%20is%20charged>. [↑](#footnote-ref-4)