



Full Name : _____ Student ID: _____

Grade Table (for Lecturer use only)

Question	Points	Score
1	5	
2	10	
3	15	
4	40	
5	20	
6	10	
Total:	100	

Instructions for Online Midterm Exam

Welcome to the midterm exam of EEE356-Data Analytics and good luck!

- The midterm exam will be conducted between 09:15 and 12:00. The system will be closed at 12:00. Students must finalise the exam by saving before 12:00.
- The students are also responsible to have a fully-charged laptop computer and 3G internet connection against an interruption in the electricity or internet services. Please note that, allowed number of entries during the exam is determined as 3 due to the internet disconnections.
- This is a closed-book exam that means students are not allowed to take notes, books, or any other reference material into the exam. Students need to rely entirely on their memory to answer questions.
- The exam must be taken completely alone. Showing it or discussing it with anybody is forbidden, including (but not limited to) the other students in the course in current or previous years. Absolutely no communication is allowed between or among students.
- An incorrect answer to a question is awarded no marks with no consideration of any partial credit. Therefore, no partial credit will be given.
- Please sign the below Honour Code statement.

In recognition of and in the spirit of the above rules which constitute Adana Alparslan Türker Science and Technology University Honour Code, I certify that I will neither give nor receive unpermitted aid on this examination.

Signature: _____



1. **(5 points)** Match the name of Tidyverse packages with the appropriate descriptions.

- | | |
|------------------|-----------------------|
| 1. dplyr () | a. Data Visualisation |
| 2. tidyr () | b. Data Import |
| 3. ggplot2 () | c. Data Manipulation |
| 4. readr () | d. Date and Time |
| 5. lubridate () | e. Data Organising |

2. Answer the following questions.

(a) **(5 points)** Define the term 'Data Analytics' with your own words.

(b) **(5 points)** Explain one of data analytics' methodologies in details.



3. (15 points) Develop an R function named as *findroot* that firstly requests coefficients of a quadratic function ($ax^2 + bx + c = 0$) from the user, then calculates Δ ($\Delta = b^2 - 4ac$) and finally computes the roots $x_{1,2}$ ($x_{1,2} = \frac{-b \pm \sqrt{\Delta}}{2a}$).

Hint:

- If $\Delta > 0$, there shall be two real roots x_1 and x_2 ;
- Else if $\Delta = 0$, there shall be only one real root owing to $x_1 = x_2$;
- Else, there shall be no real root.

4. (40 points) Download 'Daily Electricity Price and Demand Data' from Kaggle data sets and save 'complete_dataset.csv' file to your desktop folder. Import the data set into RStudio as a tibble named as 'Data' by emphasising corresponding R codes for each step and check if 'Data' tibble seems at RStudio workspace and has 2,106 observations and 14 variables.

After proper data wrangling, draw a graph by using `ggplot2::geom_path` function with `turquoise*` colour according to the followings:

- x-axis shall demonstrate dates from March 2020 to the end date of the data set and be labelled as 'Record Interval (from March 2020)',
- y-axis shall indicate demand whilst rainfall is greater than zero and be named as 'Total Daily Electricity Demand (GWh)',
- Nonetheless `ggplot2::facet_wrap` function shall be deployed in order to divide the final visual output in two columns by means of representing 7 days of a week as 1 for Monday, . . . , and 7 for Sunday.

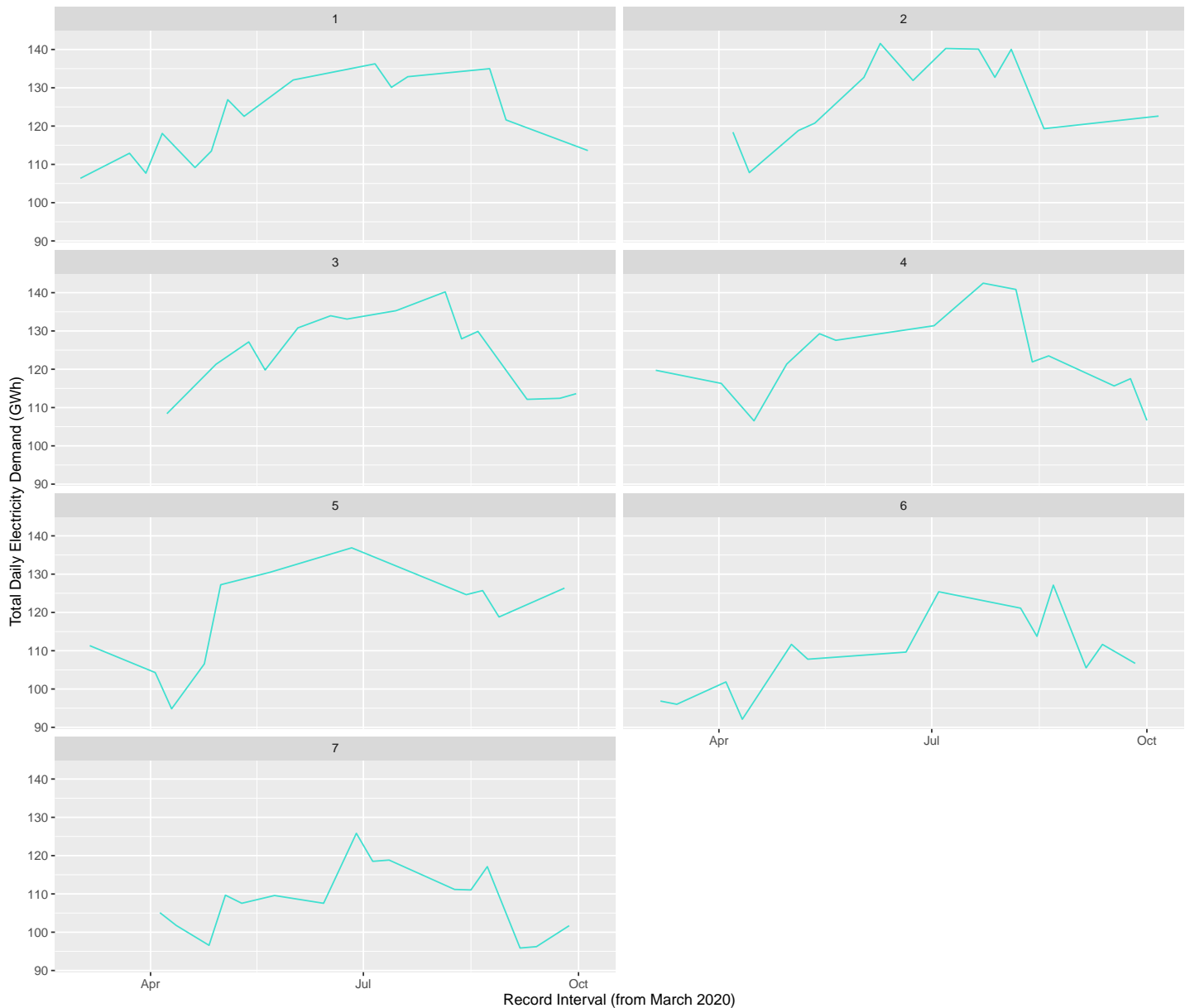


Figure 1: The visual output



The visual output can be viewed by using the below link via downloading a copy:

- <https://tinyurl.com/EEE356ExamMidterm>

**According to Wikipedia, turquoise is a blue/green color, based on the gem of the same name. The word turquoise comes from the French for 'Turkish', as the gem was originally imported from Turkey. The first recorded use of turquoise as a colour name in English was in 1573.*



5. Use 'flights' data set under 'nycflights13' package and respond the following questions by emphasising corresponding R codes:

(a) **(5 points)** Which function shall be used for a data set to access the class, size, variable names, variable data types, and several first observations belonging to each variable of the data set?

Answer: _____

(b) **(5 points)** How many flights were realised on 19th April 2013 from the John F. Kennedy (JFK) International Airport located in the city of New York?

Answer: _____

(c) **(5 points)** What was the latest flight with respect to scheduled departure time on that day in the JFK airport?

Answer: _____

(d) **(5 points)** What was the average departure delay for the flights realised on that day in the JFK airport? Please do not take negative and zero delay values into account.

Answer: _____



6. (10 points) Republic of Turkey Ministry of National Education provides scholarship to graduates for post-graduate educations in foreign reputable institutes on condition that scholars have to work in a predetermined Turkish State or Educational Institution as double time as their study periods in a foreign country. Consider such scholar started his/her postgraduate education on 20th September 2012 and completed his/her education on 18th April 2021. Assume that the scholar started to work in a Turkish State University on 19th April 2021. Calculate by emphasising R codes in order to find the closest date that the scholar can resign without paying any compensation.

Answer: _____