The International Ski and Snowboard Federation (FIS), organizes and records a series of events such as the World Cup which exists for all disciplines of FIS such as Nordic Skiing. World Cup races occur every weekend between January and March, with events hosted in different countries. These events consist of multiple different races, held in different disciplines and distances. Nordic skiing is split up into two different techniques, classic and skate, all race weekends have a mixture of skate and classic technique races. The race in this dataset is a women’s' 10k classic race set in Trondheim, Norway. In long distance pieces like 10ks there is often a lot of discussion about pacing. Racers' times are recorded at different distance intervals throughout the course. It is interesting to look at how well racers' times at different distances predict their rank at the end of the race. The first time is most interesting to see with regard to the final rank, to see if the same racers who have top times in the first distance have good ranks at the end of the race. The following questions are based on a dataset from a Nordic Ski race in Trondheim, Norway on the 17th December 2023. The race is one of many World Cup events held by the International Ski Federation throughout the months of December through March. This race in particular was a 10-kilometer women’s race in the classic technique, with 61 women enrolled. The course for this race consisted of two 5-kilometer loops. In the original data set, there are 4 different distance intervals taken at 1.3, 4.3, 7.5, and 10k respectively.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Time1 | The skier’s time after the first distance of 1.3k. |
| Rank4 | The skier’s rank at the finish of the race. |

1. Fit the following models and create scatter plots with regression lines for each of them. Record their R2 values and comment on the linearity.
   * 1. Linearity:
     2. R2:
     3. Linearity:
     4. R2:
     5. Linearity:
     6. R2:
     7. Linearity:
     8. R2:
2. Which one of the models from Question 1 do you think is best? Why?
3. Interpret the slope coefficient for the model you just chose in Question 2. Remember to contextualize.
4. If a skier is ranked #14 after the first split, what rank would you predict for her finish?
5. Interpret the slope coefficient for the best of these models to predict Time4. Remember to contextualize.
6. If a skier had a time of 5.00 minutes after the first split, what time would you predict for her finish?
7. Can it be concluded that the time and/or rank of a skier after the first distance interval has a substantial impact on their finishing rank and/or time? Explain your answer.