Skills Assessment - Predictive Modeling & Probability

1) Use the attached 'PredictiveModelingAssessmentData.csv' to develop a model, predicting response y from inputs x1 and x2. Assume the file contains no data errors. Your only goal is to minimize the prediction error measured by RMSE. Once your final model is trained, please generate predictions for 'TestData.csv'. Your predictions on the test set will be scored for accuracy. Save and return predictions on the test set as 'TestDataPredictions.csv'. Please share your code. Assume even minor improvements in accuracy can provide a significant competitive advantage. Hint: the irreducible error measured by RMSE is known to be 1.

2) The NBA draft lottery is an event in which the teams who had missed the playoffs that previous year participate in a lottery process to determine the draft order. The NBA recently implemented changes for a new system starting in 2019 but some believe further modifications to the process are necessary. For details on the process see: https://en.wikipedia.org/wiki/NBA_draft_lottery#Process

Assume that the NBA lottery odds and structure are further modified as follows:

i. Rather than deciding the top three *(old system)* or four *(new system)* picks by the lottery, the top five picks are decided by the lottery
ii. The lottery is expanded from 14 teams to 16 teams (in addition to the 14 non-playoff teams, the lottery now includes 2 playoff teams)
iii. The possible 1000 lottery combinations are redistributed as follows (use right column only):

Distribution of Chances (out of 1000)

Seed	***Old System***	***New System***	**Hypothetical**
1	250	140	114
2	199	140	113
3	156	140	112
4	119	125	111
5	88	105	99
6	63	90	89
7	43	75	79
8	28	60	69
9	17	45	59
10	11	30	49
11	8	20	39
12	7	15	29
13	6	10	19
14	5	5	9
15	-	-	6
16	-	-	4

Please complete the table of conditional probabilities below (the probability of each seed getting each pick) and save results as 'DraftProbabilityTable.csv'. Please share any code used to determine probabilities.
Example: P(1st pick | worst record) = 0.114

Seed	Chances	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th
1	114	0.114															
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
16	4	0.004															

3) Use the data in the attached 'acc1819.db' SQLite database file containing a table of game information and a table of box-score information on ACC games from the 2018-2019 NCAAB season to develop a predictive ranking of ACC teams. The final ranking should be predictive in the sense that a team ranked better according to your ranking would be projected to beat a worse-ranked team (at a neutral site). Please rank all ACC teams and return the ranking (best team 1 to worst team 15) in a file with the title 'ACCRankings1819.csv'. Creating a predictive rating for each team is optional, but if used, please include this information as a column in the same csv file. The attached 'ExampleRankings.csv' file gives an example of the desired output format using the real 2018-2019 ACC standings (and conference win percentage as the rating). Please share any code used to develop your rankings, as the purpose of this open-ended question is to see your thought process through a realistic sports modeling problem.

If you are unfamiliar with using SQLite, you may find the references below helpful:

i. For python: https://docs.python.org/3/library/sqlite3.html
ii. For R: https://db.rstudio.com/databases/sqlite/
iii. For viewing the file: https://sqlitestudio.pl/