ORIE 3120 Prelim 1

March 3, 2020

- The exam is 1.5 hours long.
- You may use one (8.5×11) sheet of notes (back and front).
- No computers, calculators, phones, or other electronic devices allowed.
- Use the space after each question for your answer. The backs of the exam pages may be used for scratchwork but will not be graded.

Name: _____
NetID: _____

Honor code. Write the following statement, and sign your name below:

I certify that all work on this exam is my own.

Honor code: _____
Signature: _____

1. *Spotify.*

Spotify Global Top 50 is ranked by the number of streams. We have the data for February 20, 2020 and February 19, 2020 in tables named Feb20 and Feb19, respectively. The following are the first 10 data entries from each table.

February 19:

	Position	Title	Artist	Streams
1	1	Blinding Lights	The Weeknd	5983876
2	2	Dance Monkey	Tones and I	5957301
3	3	The Box	Roddy Ricch	5749543
4	4	Intentions (feat. Quavo)	Justin Bieber	4833667
5	5	Don't Start Now	Dua Lipa	4633763
6	6	No Time To Die	Billie Eilish	4467091
7	7	Yummy	Justin Bieber	4037354
8	8	Tusa	KAROL G	4002459
9	9	Falling	Trevor Daniel	3871618
10	10	ROXANNE	Arizona Zervas	3714629

February 20:

	Position	Title	Artist	Streams
1	1	Blinding Lights	The Weeknd	6127151
2	2	Dance Monkey	Tones and I	5980798
3	3	The Box	Roddy Ricch	5643872
4	4	Intentions (feat. Quavo)	Justin Bieber	4753045
5	5	Don't Start Now	Dua Lipa	4601096
6	6	No Time To Die	Billie Eilish	4231654
7	7	After Hours	The Weeknd	4152572
8	8	Tusa	KAROL G	4067362
9	9	Yummy	Justin Bieber	3991664
10	10	Falling	Trevor Daniel	3821073

You may assume that each title appears only once in the chart (each track has a unique title), but there may be artists with multiple songs in the same chart.

(a) (5 points) You want to compare how the ranking changed from February 19 to February 20 for each song that was ranked on February 19. Create a query that has

- Column 1: February 19th ranking
- Column 2: February 20th ranking (or NULL if the song was not ranked on February 20)
- Column 3: Change in rankings

$$\begin{cases} 0 & \text{if the ranking did not change} \\ n & \text{if the ranking went up } n \text{ positions} \\ -n & \text{if the ranking went down n positions} \end{cases}$$

  (or NULL if the song was not ranked on February 20)
- Column 4: Title
- Column 5: Artist

The first 10 entries in the resulting view should look something like this:

	Feb19Position	Feb20Position	Difference	Title	Artist
1	1	1	0	Blinding Lights	The Weeknd
2	2	2	0	Dance Monkey	Tones and I
3	3	3	0	The Box	Roddy Ricch
4	4	4	0	Intentions (feat. Quavo)	Justin Bieber
5	5	5	0	Don't Start Now	Dua Lipa
6	6	6	0	No Time To Die	Billie Eilish
7	7	9	-2	Yummy	Justin Bieber
8	8	8	0	Tusa	KAROL G
9	9	10	-1	Falling	Trevor Daniel
10	10	11	-1	ROXANNE	Arizona Zervas

(b) (5 points) You want to see the songs for which the number of streams changed by less than 10,000. Create a query that has

- Column 1: Number of streams on February 19th
- Column 2: Number of streams on February 20th
- Column 3: Difference in number of streams on February 19th vs 20th
- Column 4: Title
- Column 5: Artist

ordered in descending order of the difference in number of streams. The resulting view should look something like this:

	Feb19streams	Feb20streams	Difference	Title	Artist
1	2603655	2596922	6733	bad guy	Billie Eilish
2	1822026	1817361	4665	Vete	Bad Bunny
3	2461242	2460221	1021	Adore You	Harry Styles
4	1568203	1567491	712	Fantasias	Rauw Alejandro
5	2058104	2057519	585	Physical	Dua Lipa
6	1890286	1891515	-1229	Lose Control	MEDUZA
7	2243397	2245634	-2237	RITMO (Bad Boys For Life)	Black Eyed Peas
8	2340900	2345883	-4983	Señorita	Shawn Mendes
9	2658748	2665035	-6287	Ignorantes	Bad Bunny
10	2100261	2110238	-9977	Blueberry Faygo	Lil Mosey

Hint: The absolute value of a number x is abs(x) in SQL

2. *Majors.* We have two databases, Students and Majors. The Students table has four columns Student ID, Name, Major and Year. You can assume each StudentID appears exactly once in the Students table (names may repeat) and that each student has exactly one major.

   The first three columns of the Students table are below.

StudentID	Name	Major	Year
1	James Smith	3	2
2	Maria Ramirez	2	2
3	Sarah Jones	7	3

   The Majors table has three columns, MajorID, MajorName, and NumberOfFaculty, which gives the number of faculty associated with the major. The first four columns of the Majors table are below.

MajorID	MajorName	NumberOfFaculty
1	American History	25
2	Computer Science	52
3	Math	23
4	Spanish	34

   (a) (5 points) You want to know which majors have the most students. Write a query that has:

   - Column 1: MajorID
   - Column 2: MajorName
   - Column 3: NumberofStudents

   The results should be in descending order of the number of students. Ensure all majors appear in the table.

(b) (5 points) You would also like to know the student to faculty ratio. Write a query, or explain how you would change to the previous query, to produce a column which gives the student to faculty ratio for each major. Call your column StudentFacultyRatio.

(c) (5 points) Finally, you want a list of all the second-year students' names and their majors. Write a query that has:

- Column 1: StudentName
- Column 2: MajorName

Order your results alphabetically by StudentName

3. *Piazza awards.* As a way to incentivize students to make good Piazza posts, the ORIE 3120 staff have decided to establish the "Piazza Awards" for the students with the most good posts on Piazza, with 1st place, 2nd place and 3rd place awards being given out. In fact, we want to mail out the awards this week — but the problem is that all we have is a single database with two tables, Posts and Students.

   Here is a portion of the Posts table:

	PostID	StudentID	GoodPost
1	1	1	TRUE
2	2	1	FALSE
3	3	2	TRUE
4	4	1	FALSE
5	5	2	TRUE
6	6	3	TRUE
7	7	3	TRUE
8	8	3	TRUE
9	9	2	FALSE
10	10	4	FALSE

   The third column, GoodPost, records whether the Piazza post was good as judged by your TAs.

   And here is a portion of the Students table:

	StudentID	StudentName	Address
1	1	John	123 Sesame St
2	2	Sophie	101 Ithaca Rd
3	3	Pierre	92 College Ave
4	4	Maria	24 Lake St

   We want to determine where to mail each of the awards, using SQL queries.

(a) (5 points) Create a view, GoodPostsCount, that shows the number of good posts each student has made. It should have two columns:

- StudentID
- NumGoodPosts, which records the number of good posts made by the student.

For example, the table might look like this:

	StudentID	NumGoodPosts
1	3	3
2	2	2
3	1	1

(Order of rows does not matter.)

(b) (5 points) Create a view, PairwiseComparisons, that has one row for each pair of student IDs, and four columns:

- PostedMoreID is the StudentID of the student who posted more good piazza posts
- PostedLessID is the StudentID of the student who posted fewer good piazza posts
- MorePosts is the number of good posts made by PostedMoreID
- LessPosts is the number of good posts made by PostedLessID

Your result should look like this:

	PostedMoreID	MorePosts	PostedLessID	FewerGoodPosts
1	3	3	3	3
2	3	3	2	2
3	3	3	1	1
4	2	2	2	2
5	2	2	1	1
6	1	1	1	1

Notice that there is one row comparing each student with themselves. You may assume that there are no ties: no two students have made exactly the same number of good piazza posts.

(c) (5 points) Using the results of the previous view, create a table with three columns:

- StudentID
- Address
- Rank: the student with the most good posts should have value 1, the student with the second most good posts should have value 2, etc.

and three rows corresponding to the top three ranked students, ordered by rank. Your result should look like this:

	StudentID	Address	Rank
1	3	92 College Ave	1
2	2	101 Ithaca Rd	2
3	1	123 Sesame St	3

4. *PearPhone.* You lead a large tech company that produces phones called the PearPhone. Fortunately, the market is quite predictable and you have a constant demand rate over time. This demand rate is 18 per month. Suppose you need to reorder chips to produce your phone and the lead time is 1 week. Here, we will operate under EOQ conditions and we let assembly cost be virtually \$0 because it will be built by robots in a very quick time efficient manner. The market price of each phone sold is \$500 whereas, the chips required to build the phone cost \$100 and the fixed cost of a shipment of chips is \$100. Finally, holding chips required for a single PearPhone in your warehouse costs \$1 per week. For this question, you may assume every month is 4 weeks.

   Using this information, find the following.

   (a) (3 points) What is the optimal order quantity?

   (b) (3 points) What is the average profit made every month?

   (c) (3 points) What is the inventory level when an order is made?

(d) (3 points) How many orders will be placed in one year?

5. *Newsvendor.* Consider the Newsvendor Problem, where we are selling newspapers, which we buy for \$1 and sell for \$1.50. If we have more newspapers than the demand the remaining newspapers will be recycled, and we get \$0 for each of these.

   Assume that the demand is a random variable with cdf $F(\cdot)$. Assume that this cdf is invertible, and denote this inverse by $F^{-1}(\cdot)$.

   We modeled the cost of newsvendor by thinking about the "opportunity costs". For each newspaper that could have been sold, the newsvendor is missing out on some potential profit (this was the "underage cost"), and for each newspaper that was not sold, the newsvendor is paying a cost (this was the "overage cost").

   (a) (3 points) Give the value of the per unit overage cost and the per unit underage cost for the example above.

   (b) (3 points) Give an expression for the total cost if the demand is x (a number, not a random variable), and the number of newspapers at the start of the day is Q.

(c) (3 points) Give an expression for the number of newspapers to buy that minimizes the expected cost.

We will now consider the "Fish Fragrant Newsvendor Problem": In this problem, we are selling newspapers, which we buy for \$1 and sell for \$1.50. Now however, when we have 20 or less left at the end of the day, we can sell them to the local fish restaurant as wrappers for their fish dinners for \$1 per newspaper. If we have more left, we can sell 20 to the local fish restaurant for \$1 per newspaper, but have to bring the remaining newspapers to the recycling, and we get \$0 for each of those.

We again want to decide the optimal number of newspapers to buy at the start of the day, Q. As in the normal Newsvendor Problem, we will think about minimizing the opportunity cost.

(d) (3 points) Give an expression for the cost when we decide to start off with Q newspapers and the demand is x (a number, not a random variable) — we'll refer to this as $G(Q, x)$.

(Hint: it may help you to answer these questions, to derive the expression. These answers will not be graded.

1. Suppose x is the demand for newspapers in a day (a number, not a random variable) and suppose $x \geq Q + 1$. How much more profit could we have made

if we decided to buy $Q+1$ newspapers at the start of the day, instead of just Q newspapers?

2. Suppose x is the demand for newspapers in a day (a number, not a random variable) and suppose $Q \geq x + 20$. What would have happened to our profit if we decided to buy $Q+1$ newspapers at the start of the day, instead of just Q newspapers?
3. Suppose x is the demand for newspapers in a day (a number, not a random variable) and suppose $x \leq Q \leq x+19$ (in other words, we have ordered at least as many newspapers as the demand, and there are 19 or fewer newspapers left at the end of the day). What would have happened to our profit if we decided to buy $Q+1$ newspapers at the start of the day, instead of just Q newspapers?)

(e) (3 points) Find the derivative of the expression $G(Q, x)$ with respect to Q.

If you did not get an expression for the cost, you can use $\frac{d}{dQ}G(Q,x) = 3\mathbb{1}_{\{(Q-x)^+\}} + 5\mathbb{1}_{\{(x-Q-10)^+\}}$ from here on, where $\mathbb{1}_{\{\cdot\}}$ is the indicator function as you have seen in class.

(f) (3 points) Now let D be a random variable denoting the demand. Give an expression for the expected cost when we decide to start the day with Q newspapers.

(g) (3 points) Take the derivative of the expected cost with respect to Q assuming that the derivative of this expectation is equal to the expectation of the derivative that you found in e.

Simplify the expression as much as possible.

(h) (3 points) Find an expression that must be true for the optimal number of newspapers that we should buy at the start of the day based on the previous question.

If you did not find a derivative in the previous part, you can pretend that the derivative of the expectation equals $F(Q) - 2(1 - F(Q + 3))$, where $F(\cdot)$ is the cdf of the random variable D, and use this in this question.

(i) (3 points) Explain why you can use bisection search to find the optimal number of newspapers (approximately).

Question	Points	Score
1	10	
2	15	
3	15	
4	12	
5	27	
Total:	79	