# **CIS 3252 Final Exam Answer Template**

# **Part 1. Descriptive Analytics (20 points)**

**Pomona Temperature Data**

* 1. **Calculate descriptive statistics for this data.**

**Table

Description automatically generated**

<https://public.tableau.com/app/profile/rebecca1664/viz/FINAL1_16705507094570/Sheet1?publish=yes>

* 1. **Create an appropriate chart that displays this data.**

Chart, bar chart

Description automatically generated

* 1. **Describe what your descriptive statistics say about the data.**

The descriptive statistics show the high temperatures on the top and the low temperatures on the bottom for each month of the year. They are color coordinated to differentiate the months of the year as well. The months are along the x-axis due to them being the independent variable and the temperature is along the y axis due to them being dependent variables. The months with the highest temperatures in Pomona are July and August. The months with the lowest temperatures in Pomona are December and January.

* 1. **Is there other data that could be shown that might be useful to someone visiting Pomona?**

Other data that could be useful to someone visiting Pomona could be the total amount of rainfall for each month of he year. I

1. **In the final dataset spreadsheet, use the Product Sales Data for this question. The data shows the sales totals for two Sales People for different cities near Pomona. You’ve been asked by a local business to look at the sales data and make recommendations.** 
   1. **List what questions you might be able to answer with this data and why that is important to your client.**

How many sales did each sales person make?

In what city did the salesperson make the most sales?

* 1. **Create a box plot for the data for Salesperson 1 and Salesperson 2.**

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

* 1. **Describe what the box plot tells you.**

The box plot tells you the median sales, the range of sales, as well as the highest and lowest sales of each salesperson. Based on the box plots, salesperson 1 is making more sales.

* 1. **As a manager, what decisions can you potentially make from this data?**

Decisions that a manager could potentially make from this data is which salesperson to hire or not hire.

* 1. **What data could be collected to make the decision more valuable?**

Data that can be collected to make the decision more valuable can be what the person was selling, when did they make sales, how long did it take to make the sales, and what is their average monthly sales.

**Part 2. Predictive Analytics (30 points)**

1. **Regression – In the final dataset spreadsheet, use the Social Media Screen Time data for this question. Investigate the regression relationship between the weekly screen time of TikTok vs. Facebook.**
   1. **Determine Outliers – What decisions did you make in removing outliers from the data? Show a picture of your data before and after the removal of outliers.**

**Chart

Description automatically generatedBEFORE**

**Chart

Description automatically generated AFTER**

* 1. **What is the linear relationship between TikTok and FaceBook?**

Chart, scatter chart

Description automatically generated

* 1. **What is the equation for the regression line? (Excel will do this for you)**

y = 0.0248x + 3.6044

* 1. **Calculate the outcome of TikTok equivalent to 1.3, 5.7, and 9.1 based on your regression line.**

1.3 = 3.6118

5.7= 3.7458

9.1= 3.8301

* 1. **What is your R-squared value? What does that mean?**

R² = 0.0462, with facebook variable, you can only explain 4% of this data

* 1. **Is TikTok use a good predictor of the amount of time spent on FaceBook? Give a paragraph answer describing why. What does the data tell you?**

Tik Tok use is not a good predictor of the amount of time spent on Facebook. This is in part due to the low r-squared value. R-squared essentially measures the correlation between the two variables. Since this r-squared equals 0.0462, that means that there is only a 4% correlation between the hours spent on Tik Tok and the hours spent on Facebook by an average user.

1. **K-Means (Cluster Analysis) – Tableau**
   1. **Import the data into Tableau and perform a cluster analysis.**

Chart, application, bar chart

Description automatically generated

* 1. **Describe what that cluster analysis tells you.**

Cluster analysis shows us groups, or clusters, in the data that were not visible when viewing raw data.

# **Part 3 Visual Dashboard and Data Story in Tableau (50 points)**

* 1. **Problem Description**
     1. **What data did you use? Provide a link and describe why you selected it.**

I selected 2015 flight delays and cancellations data set. I selected this dataset because flight cancellations and even delays can be a big disruption in people’s lives. Noticing if there’s patters of this occurring in specific airlines would be useful information for travelers to know.

<https://www.kaggle.com/datasets/usdot/flight-delays>

* + 1. **Who is the audience for your analysis and what questions might they be interested in? 4 Questions.**

The audience for my analysis would be travelers in 2016. They would want to see data from the previous year to determine which airline would be the best to take for their trip. They would want to know which airline had the most cancellations and the most delays for the previous year, which would be 2015.

* + 1. **Are there any ethical issues with this data set?**

**T**here is no ethical issues with this data set since it is just recording factual events without any bias.

* 1. **Create a Dashboard in Tableau that shows answers to the questions you developed in part a. Give an answer to each question.**

The airline with the most delays is the airline labeled ‘WN’ which is to represent Southwest Airlines. The airline with the most cancellations is labeled ‘MQ’ which represents Envoy airlines.

* 1. **Create a Story that uses the Setup/Conflict/Resolution approach to data story telling. Make it engaging, informative, and lovely.**

A man was away on business and had to make it home in time for the holidays to meet his wife and kids. He noticed his flight had been delayed and needed to make a split decision, so he pulled up the data set to make the right decision. He made the right decision because right as he was boarding his new flight, he heard the announcements of the cancelled flights and knew the right decision was made so that he could make it home in time for the holidays

* 1. **How does this dataset build actionable insights?**

The data set builds actionable insights because it helps travelers make decisions as to make it to their destination in time

* 1. **Provide a link to your dashboard and data story.**

<https://public.tableau.com/authoring/RebeccaRojas_part3/Story1#1>