DATA-321
Exam I

Tableau Question 1

`Goggle Play Apps Store` dataset contains information related to apps downloads. Each app (row) has values for category, rating, size, and more. The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market. Here is a description of each of the fields/variables in the `Goggle Play Apps Store` dataset:

- `App:` application name.

- `Category:` category the app belong to.

- `Rating:` overall user rating of the app.

- `Reviews:` number of user reviews for the app.

- `Size:` the size of the app.

- `Installs:` number of user downloads/installs for the app.

- `Type:` paid or free.

- `Price:` price of the app.

- `Content Rating:` age group the app is target at.

- `Genres:` an app can belong to multiple genres (apart from its main category). For example, a musical family game will belong to music, game, family genres.

- `Last Updated:` date the app was last updated on play store.

- `Current Ver:` current version of the app available on play store.

- `Android Ver:` minimum required android version.

1. (3 points) Load `googleplaystore.csv` file to tableau.

2. (4 points) What is the most popular app category? What is the least popular app category?

3. (4 points) What are the top three app genres?

4. (6 points) Create an appropriate chart that shows the number of app updates over time based on `Content Rating`. Describe your chart.

5. (6 points) What is the app name with most download from the `Content Ratings` equal to `Everyone`?

6. (8 points) Create an appropriate visualization that compares the distribution of rating for each `Content Rating`. Comment on your chart.

7. (8 points) Assuming that you are working for a finance company. How would use this data to decide what day of the week would you release the app? Explain your answer with an appropriate chart.

Tableau Question 2

The `police_killings.csv` data file contains the data behind the story
[Where Police Have Killed Americans in 2015](). The data was linked using entries from the
Guardian's database on police killings to census data from the American Community Survey.
The Guardian data was downloaded on June 2, 2015. Census data was calculated at the tract
level from the 2015 5-year American Community Survey using the tables `S0601 (demographics)`,
`S1901 (tract-level income and poverty)`, `S1701 (employment and education)` and `DP03
(county-level income)`. The following table shows a brief description of each of fields:

Name	Description	Source
Name	Name of the deceased	Guardian
Age	Age of the deceased	Guardian
Gender	Gender of the deceased	Guardian
Raceethnicity	Race/Ethnicity of the deceased	Guardian
Month	Month of the killing	Guardian
Day	Day of the incident	Guardian
Year	Year of the incident	Guardian
Streetaddress	Address where incident occurred	Guardian
City	City where incident occurred	Guardian
State	State where incident occurred	Guardian
Latitude	Latitude, geocode from address	
Longitude	Longitude, geocode from address	
State Fp	State FIPS code	Census
County Fp	County FIPS code	Census
Tract Ce	Tract ID code	Census
Geo Id	Combined tract ID code	
County Id	Combined county ID code	
Namelsad	Tract description	Census
Lawenforcementagency	Agency involved in incident	Guadian
Caused	Cause of death	Guardian
Armed	Whether deceased was armed	Guardian
Pop	Tract population	Census
Share White	Share of pop that is white	Census
Share Black	Share of pop that is African-American	Census
Share Hispanic	Share of pop that is Hispanic	Census
P Income	Tract-level median personal income	Census
H Income	Tract-level median household income	Census
County Income	County-level median household income	Census
Comp Income	`H Income / County Income`	Calculated from Census
County Bucket	Household income, quintile within county	Calculated from Census
Nat Bucket	Household income, quintile nationally	Calculated from Census
Pov	Tract-level poverty rate	Census
Urate	Tract-level unemployment rate	Calculated from Census
College	Share of 25+ pop with BA or higher	Calculated from Census

Notes regarding income calculations:

- All income fields are in inflation-adjusted 2013 dollars.

- `Comp Income` is simply tract-level median household income as a share of county-level median household income.

- `County Bucket` provides where the tract's median household income falls in the distribution (by quintile) of all tracts in the county. (1 indicates a tract falls in the poorest 20% of tracts within the county). Distribution is not weighted by population.

- `Nat Bucket` is the same as `County Bucket` but for all US counties.

Using the `police_killings.csv` data file answer the following:

1. (3 points) Load the `police_killings.csv` data file to tableau.

2. (5 points) Create an appropriate chart that shows the number of police killings per state. What state had the highest number of police killings in 2015? What state had the lowest number of police killings in 2015?

3. (5 points) Create an appropriate chart that shows the Race/Ethnicity of the deceased. What is the most frequent Race/Ethnicity? What is the least frequent Race/Ethnicity?

4. (5 points) Add your chart from question 3 to your chart in question 2 under tooltip. Is there a relationship between state and Race/Ethnicity of the deceased? Explain.

5. (5 points) What month has the higher number of police killings? Create an appropriate chart that shows the number of police killing cases per month.

6. (5 points) Is there a relationship between cause of death and whether the deceased was armed? Create an appropriate visualization that explore that relationship.

7. (3 points) `Urate` represents the unemployment rate. In the data file, it is store as a number. Change the `Urate` number to percentage. For example, 15 should be 15%

8. (7 points) A common belief is that there is a relationship between crime and unemployment rate. That is, as the unemployment rate increases, crime increases as well. That said, let's explore if there is some sort of relationship between police killings and unemployment rate. Create an appropriate visualization that shows the number of crimes per state/city and unemployment rate. Comment on your chart.

9. (5 points) Create a new calculated field, call it `Pov Percentage`. Notice that `Pov Percentage` is defined as `Pov / 100`. Make sure `Pov Percentage` is display as a percentage number.

10. (7 points) Does cities with higher poverty rates have more police killings? Create an appropriate visualization that shows the number of crimes per state/city and poverty rate. Make sure you use `Pov Percentage` from question 9. Comment on your chart.

11. (3 points) `College` represents the share of 25+ pop with BA or higher education degree. In the data file, it is store as a number. Change the `College` number to percentage. For example, 0.1685 should be 16.85%

12. (5 points) Create a parameter, call it `Education Percentage`. Set the data type equal to float, current value 0.05, and range from 0.01 to 0.3 and step size 0.01.

13. (5 points) Create a new calculated field, call it `Filter Education`. Notice that this new field will filter cities with `College` greater than the value of `Education Percentage`.

14. (7 points) Create an appropriate visualization that shows the number of crimes per state/city with `College` greater than `Education Percentage`. What city has the highest number of police killings with `College` greater than 25%?

15. (5 points) Create a chart that shows the cause of death. Add it as tooltip to your chart in question 14. What the most common cause of death in Tulsa, OK?

16. (10 points) Create a visualization that show that state/cities with police killings cases for a given name. What cities had police killing cases with name `chris`?

Conceptual Tableau Questions

1. (4 points) What steps should be taken to change a vertical bar chart to horizontal?

   (a) Drag the measure to the columns shelf and the dimension to the row shelf.
   (b) Drag the measure to the rows shelf and the dimension to the column shelf.
   (c) Drag the measure to both the row shelf and the columns shelf.
   (d) Drag the dimension to both the rows shelf and the column shelf.

2. (4 points) How to create highlight table or heat map in Tableau:

   (a) Drag one or more dimensions on the columns shelf and one or more dimensions on the rows shelf. Then, place a measure of interest on the color shelf. Finally, select square as the mark type.
   (b) Drag one or more dimensions on the columns shelf. Then, select square as the mark type. Finally, place a measure of interest on the color shelf.
   (c) Drag one or more dimensions on the rows shelf. Then, select Square as the mark type. Finally, place a measure of interest on the color shelf.
   (d) Drag one dimension on the columns shelf and one measure on the rows shelf. Finally, place a measure of interest on the color shelf.

3. (4 points) In Tableau, a discrete field is shown in which color?

   (a) Red
   (b) Blue
   (c) Green
   (d) Yellow

4. (4 points) Which of the following is **TRUE** for measures in Tableau? Select all that apply.

(a) Measures contain numeric, quantitative values.

(b) Measures contain qualitative values (such as name, or geographical data).

(c) Measures can be used to categorize, segment, and reveal the details in the data.

(d) Measures can be aggregated.

5. (4 points) Which of the following chart types requires at date field?

(a) Line chart

(b) Bar chart

(c) Pie chart

(d) Scatter plot

6. (4 points) Which of the following is a table calculation?
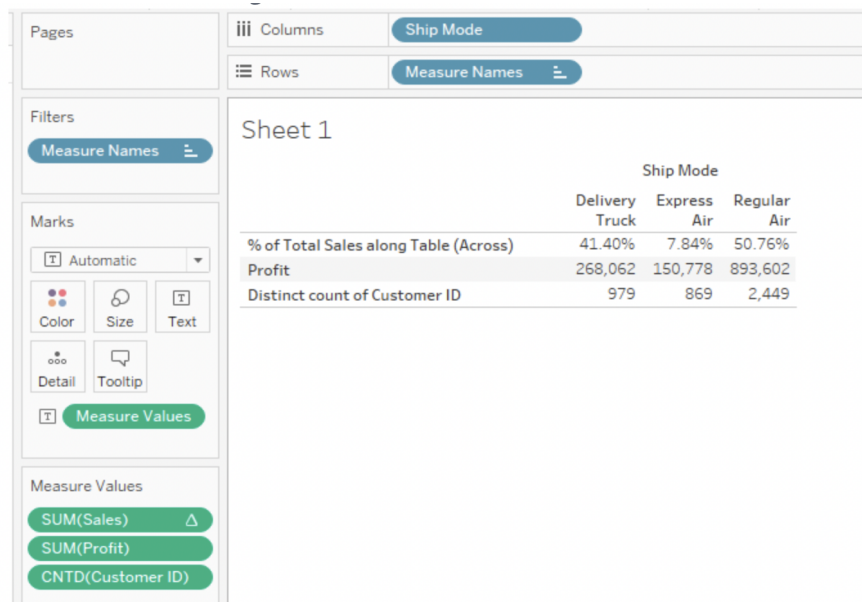
(a) Profit

(b) Distinct count of Customer ID

(c) % of Total Sales along Table (Across)

(d) None of the above

7. (4 points) Which Tableau chart would be the most appropriate for showing the relationship between two measures?

(a) Stack bar chart

(b) Combination chart

(c) Line chart

(d) Scatter plot

   (e) Histogram

   (f) Box-plot

8. (4 points) Which term best describe the below chart?
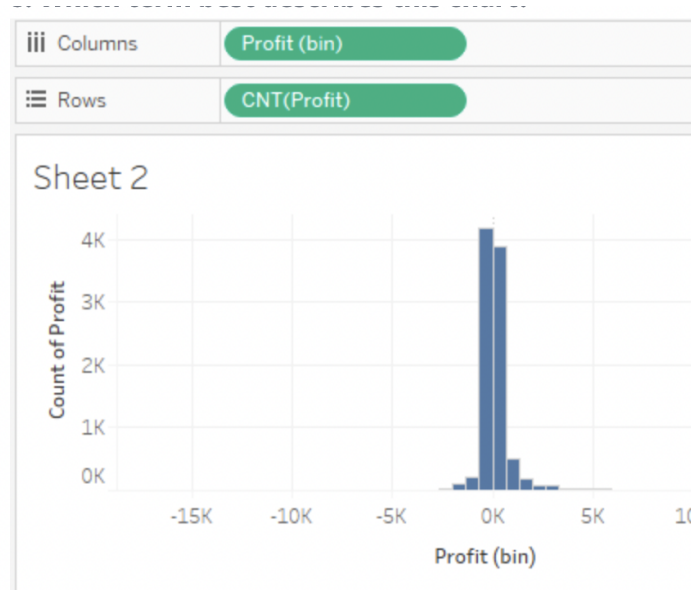
   (a) Vertical bars

   (b) Horizontal bars

   (c) Histogram

   (d) Combo chart

   (e) None of the above

9. (4 points) This type of Level of Detail expression computes total sales for the region, regardless of what dimensions are shown in the view?

   (a) {SUM([Sales])}

   (b) {FIXED [Region]: SUM([Sales])}

   (c) {ONLY [Region]: SUM([Sales])}

   (d) {EXACT [Region]: SUM([Sales])}

   (e) {INCLUDE [Region]: SUM([Sales])}

   (f) {EXCLUDE [Region]: SUM([Sales])}