

Think North Group Assessment

The Salesforce Platform

Salesforce is the world's top-rated customer relationship management (CRM) software platform. Its primary uses are sales and customer service. They offer many customized versions of the software for specific industries such as financial services and education, but they're all centered around the same primary data sets. The platform is founded upon a relational database containing data sets for things like accounts, contacts for those accounts, phone calls, emails, etc.

CRM Analytics is Salesforce's data visualization platform. It can connect to internal and external data sets. Salesforce's predictive modeling platform is Einstein Discovery, and the data sets used for the predictive models must be data sets within CRM Analytics. So, performing extract, transform, and load (ETL) within CRM Analytics is an essential skill for this position. For this assessment, we've provided you with CSV files to create the data sets that will be used as inputs. The evaluation of your data science skills requires no ETL. We're providing you with a simplified dataset that requires minimal preprocessing out of consideration for your time.

Data Science Assessment

We're using a sample of sales data for accounts in a Salesforce instance. The Annual Revenue, Average Deal Size, Top Product Total Price, and Total Revenue fields are all in US Dollars. We'll be developing a simplified predictive model for Total Revenue.

1. Please develop a model using Python, R, or both using the attached Account Sales.csv file. Your output variable is Total Revenue. Parameter tuning is not required, but showing your knowledge of it is a plus.
2. Please put comments or markdown in your code file that describes your approach to the following stages of model development and your conclusions about them, exploratory data analysis, data preparation, feature engineering, feature selection, and model selection.

ETL & Data Visualization Assessment

ETL in CRM Analytics

We'll use datasets from a Brazilian e-commerce company, Olist. Once we join them, we'll be at the order-item grain; that is, there can be multiple items for each order.

1. Create your free CRM Analytics & Einstein Discovery-Enabled Developer Edition Salesforce instance here, <https://trailhead.salesforce.com/de/promo/orgs/analytics-de>.
2. Download the attached CSV files in the Olist Sample Datasets.zip file.
3. Create a dataset for each of them in CRM Analytics. Ensure the correct data types for each column while importing them.
4. Go to the data manager and create a new recipe.
5. Join the datasets together according to the attached schema. We will not be using the geolocation, customers, or payments datasets referred to in the schema.

6. Replace the product category names with their English translations.
7. Create three derived fields.
 1. Delivery Time [Days]
 1. The time, in days, between the order purchased timestamp and the customer delivered timestamp.
 2. Delivered Late
 1. A True or False variable that indicates if the customer delivered timestamp is before or after the estimated delivery timestamp.
 3. Average Seller Review Score
 1. The average review score per seller id.
8. Save the output in a new dataset called Olist.

Data Visualization in CRM Analytics

All queries will be done on the Olist dataset. Compact-Form (queries created with the user interface), SAQL, or SQL may be used in the queries.

1. Create a new, blank dashboard.
2. Set the following layout properties.
 1. Set the columns property to 50.
 2. Set the Row Height to Fine.
 3. Set the Cell Spacing (Horizontal and Vertical) to 0.
 4. Change the Background Color Grid and Gutter properties to white.
3. Bring a container widget to the top, add a text widget, and use it to add the title Olist.
4. Beneath this container widget, drag out a filter widget and configure it to contain all global filters, including the order purchase timestamp, the order status, and the translated product category names.
5. Create a gauge chart that shows the sum of the price field across the whole dataset.
 1. Configure the settings so that there are three color bands, \$0 to \$1.7M, \$1.7M to \$3.4M, and \$3.4M to \$5M.
 2. Format the number as a currency.
 3. When you're done, it should look like the Sum of Price Gauge Chart image attached.
6. Create a gauge chart that shows the number of orders. (not the number of order items).
 1. Configure the settings so there are three color bands, 0 - 10k, 10k - 20k, and 20k - 30k.
 2. It should look like the Count of Orders Gauge Chart image attached when you're done.
7. Create a stacked column chart where each bar is a month of the order_purchase_timestamp field, and the bar segments are the Delivered Late field. The value of each bar segment is the sum of the price field.
 1. Set the properties so that the x-axis and x-axis label are showing, but the y-axis and y-axis label are not showing. Show the legend. And format the values as a currency.
8. Create a bar chart that shows the top five sellers for each product category and their rankings by the sum of price so that the seller with the highest sum of price has the lowest ranking number, 1. We want to show the sums of price, not the rankings. (Hint: This will require use of the dense_rank SAQL function, or an equivalent SQL function, and an additional projection statement.)
 1. Set the properties so that the x-axis, x-axis label, and y-axis label are showing, but the y-axis is not showing. Show the legend. And format the values as a currency.
9. Create a detail table showing the following fields, order purchase timestamp, order delivered customer timestamp, Delivery Time [Days], Delivered Late, customer_id, order status, product category name, product id, review score, review comment title, review comment message, and the seller id.

8. The labels shown on the dashboard for the axis labels and the detail table are not user-friendly. Update them in the Olist dataset's XMD file. Then refresh the dashboard to see the updated labels.

Please refer to the attached image files in the Dashboard Images.zip file for how the dashboard should look. The numerical values you obtain may be slightly different from the images.

Deliverables

Please reply with the code file showing your data science methodology and the username and password to your developer edition org, so we can log in and look at the dataset and the dashboard that you've created. When we do, we may be asked for a verification code that will be sent to the email address you used to create the Salesforce instance. So please expect us to contact you about that.

Thank you again for taking the time to complete this assessment. We hope for your success. There are some resources listed below that you might find helpful along the way.

Resources

CRM Analytics Help Documentation: https://help.salesforce.com/s/articleView?id=sf.analytics_landing_page.htm&type=5

Salesforce Analytics Query Language (SAQL) Developer Guide (Used in dashboard queries): https://developer.salesforce.com/docs/atlas.en-us.bi_dev_guide_saql.meta/bi_dev_guide_saql/bi_saql_intro.htm

EA-SQL Documentation (Used in recipes): https://help.salesforce.com/s/articleView?id=sf.bi_integrate_recipe_transformation_formula.htm&type=5