



## Activity Overview

---

In this activity, you will complete a project that showcases your ability to use Python to import, inspect, and organize data. You will also update team members through an executive summary, demonstrating your ability to organize and communicate key information.

For additional information on how to complete this activity, review the previous readings: [End-of-course project introduction](#) and [Course 2 end-of-course portfolio project overview: Waze](#).

Be sure to complete this activity before moving on. The next course item will provide you with completed exemplars to compare to your own work. You will not be able to access the exemplars until you have completed this activity.

## Scenario

---

Your team is still in the early stages of their project to develop a machine learning model to predict user churn.

Previously, you were asked to complete a project proposal by your supervisor, May Santner. Now, you have received notice that your project proposal has been approved and your team has been given access to Waze's user data. To get clear insights, the data must be inspected, organized, and prepared for analysis.

You discover two new emails in your inbox: one from May Santner, and one from your teammate, Chidi Ga. In the email, May asks for your help reviewing the data and completing a code notebook, and Chidi shares the details of the notebook. Review the emails, then follow the provided instructions to complete the PACE strategy document, the code notebook, and the executive summary.

*Note: Team member names used in this workplace scenario are fictional and are not representative of Waze.*

---

Email from May Santner, Data Analysis Manager

Subject: Help with coding notebook?

From: "May Santner," [May@waze](mailto:May@waze)

Cc: "Chidi Ga," [Chidi@waze](mailto:Chidi@waze)

Good morning!

I have a couple of updates on the user churn project. First off, the project proposal you completed has been approved. Thanks for all your great work so far. Second, I just received an email from our Senior Project Manager, Sylvester Esperanza, that our team has been given access to the Waze user data.

Before we begin the process of exploratory data analysis (EDA), we could really use your help with coding and prepping the data. During your interview, you mentioned that you worked with Python in your Google certificate program. You can draw on your Python skills for this task.

Chidi Ga (cc'd above) started a Jupyter notebook with the relevant dataset (imported). Right now, Chidi is busy finishing up a previous project. I'm sure he could use your assistance in completing the coding and setting up the notebook for the user churn project.

Chidi, do you mind sharing the details?

Best,

May Santner

Data Analysis Manager

Waze

-----  
Email from Chidi Ga, Senior Data Analyst

Subject: RE: Help with coding notebook?

From: "Chidi Ga," [Chidi@waze](mailto:Chidi@waze)

Cc: "May Santner," [May@waze](mailto:May@waze)

Nice to meet you (virtually)!

Hope you've enjoyed your first few weeks at Waze!

The project proposal you helped prepare covered the major points of this project, so I'll get right to how you can assist the team. Right now, a number of us are busy making adjustments to the machine learning model for another project, so your help is greatly appreciated!

Until we finish our previous project, there is no need to do a full EDA on our new user data. We'll get to that soon. Meanwhile, do you mind reviewing the imported data for the team? It would be fantastic if you could include a summary of the data types for each variable, where missing values exist in the data, key descriptive statistics, and anything else code-related you think is worth sharing in the notebook. I haven't had a chance to explore the data, so I really appreciate you getting an early start on this.

Thanks,

Chidi Ga

Senior Data Analyst

Waze

## Data Dictionary



This project uses a dataset called waze\_dataset.csv. It contains synthetic data created for this project in partnership with Waze. Examine each data variable gathered.

The dataset contains:

14,999 rows – each row represents one unique user

13 columns

Column name	Type	Description
ID	int	A sequential numbered index
label	obj	Binary target variable (“retained” vs “churned”) for if a user has churned anytime during the course of the month
sessions	int	The number of occurrence of a user opening the app during the month
drives	int	An occurrence of driving at least 1 km during the month
device	obj	The type of device a user starts a session with
total_sessions	float	A model estimate of the total number of sessions since a user has onboarded
n_days_after_onboarding	int	The number of days since a user signed up for the app
total_navigations_fav1	int	Total navigations since onboarding to the user’s favorite place 1
total_navigations_fav2	int	Total navigations since onboarding to the user’s favorite place 2
driven_km_drives	float	Total kilometers driven during the month
duration_minutes_drives	float	Total duration driven in minutes during the month
activity_days	int	Number of days the user opens the app during the month

driving_days	int	Number of days the user drives (at least 1 km) during the month
--------------	-----	---