

## STEPS

1. Load the *finance\_liquor\_sales* dataset in MySQL Workbench and run the necessary query to get the data for the specified time period (*SELECT \* FROM finance\_liquor\_sales WHERE DATE BETWEEN '2016-01-01' AND '2019-12-31';*)
2. Export the query result in a new csv file.
3. In PyCharm, we start by importing all the necessary libraries (Pandas for data manipulation and Matplotlib for visualizations).
4. Read the csv file that was previously extracted.
5. Drop all unnecessary columns from the data table, in order to work with a minimal dataset.
6. Aggregate the data to get the total number of bottles sold for each item in each zip code.
7. Sort the data by zip code and number of bottles sold to get the most popular item in each zip code.
8. Print the result of most popular item in each zip code.
9. Find the total sales per store and calculate the percentage of total sales for each store.
10. Sort the stores by their percentages of total sales.
11. Create a scatter plot showing the most popular item in each zip code, with colors based on the numbers of bottles sold.
12. Set the axis labels, title, ticks and colors to match the black background.
13. Add annotations for the top 8 most popular items in the plot (8 because there are 8 dots in the scatter that stand apart from the rest).
14. Display the plot
15. Similarly, create a horizontal bar plot to show percentage of sales per store number.