Python Project – '*Happy Deliveries*'

You have been employed as a Data Analyst for a food delivery business based in Ireland '*Happy Deliveries.*' The competition within the food delivery industry is fierce and the company is hoping that it can use data to help them get ahead of competitors.

You have been given **two** datasets;

- **orders**: Contains details of orders placed with *Happy Deliveries* from 2021 & 2022*.*
    1. 'order_id': unique ID of order
    2. 'order_timestamp': Timestamp for when customer placed the order.
    3. 'delivered_timestamp': Timestamp for when order was delivered.
    4. 'driver_id': ID of delivery driver.
    5. 'restaurant_id': ID of restaurant ordered from.
    6. 'cust_id': ID of customer who placed the order.
    7. 'delivery_region': County in Ireland the order was delivered.
    8. 'discount_applied': True or false whether or not the customer applied a discount to order
    9. 'discount_code': Discount code applied, note that not all orders have discounts applied.
    10. 'order_total': cost of the order **before** discount in euros (€)
    11. 'discount_pc': percentage discount
    12. 'status': The status of the delivery, whether it was successfully delivered of not.
- **customers**: Loyalty card customers of *Happy Deliveries* which contains customer details.
    1. 'id': unique ID of customer
    2. 'first_name': Customers first name.
    3. 'last_name': Customers last name.
    4. 'age': Age of customer.
    5. 'city': City in Ireland customer is based, note customers city may be different from their 'delivery_region'
    6. 'email': Email of customer.

Happy Deliveries **have provided you with the following questions to conduct your analysis:**

Reporting:

For the following **5** questions you are expected to include data visualizations:

1. Compare **monthly** 2021 sales to 2022 and determine has 'Happy Deliveries' sales grown.
2. Looking to the loyalty card holders, what is the age distribution of those customers?

3. Is there a relationship between the amount spend by a loyalty card holder and their age?
4. Is there a relationship between the amount of a payment, the age of a person **and** whether or not they used discount codes (HINT: scatterplot with 3 layers)
5. Compare the sales for 2022 across **all** regions.

Questions:

6. Christmas is coming and Happy Deliveries want to reward their high spending customers. Who are the **top 10 highest spending customers** in **2022**?
Find their ID, name and email address for the marketing department to contact them with a reward.
7. In order to keep up with the anticipated increase in sales over Christmas, Happy Deliveries want to find what are the **top 3 restaurants**, so they can allocate their resources more efficiently. What are the top 3 restaurants in terms of sales for **2022**?
Find their name, and total sales.
8. The marketing department wants to reach out to non-returning loyalty card customers from **2021** in hopes they can lower their customer churn rate. Find all the customers who are considered non-returning including their id, name and email.
(HINT: Non-returning means they have only made only one purchase)
9. The sales team want to find out if the discount code 'BLACKFRIDAY22' was as successful as last years 'BLACKFRIDAY21'. Find the total amount of sales from both discount codes.
10. What were the locations with the **lowest** cumulative sales for **2022**? Should the marketing department tailor their marketing efforts more towards this location (Perhaps other factors are responsible for example low population)

Tips to consider:

- Input errors may have resulted in errors in 'customer' dataset
- Actual payment for order needs to be calculated from the 'delivery' dataset
- Orders with the status 'CANCELLED' payments should be removed from sales since they had to be refunded.