Context

Welcome to our team – “Delivery Dron”. We are a young and ambitious team that is just starting its big journey in business, but with serious investments and a competent business plan.

If we couldn't scare you, then let's go!

We are eager to see cool analysts in our team who will take on the role of our oracles and work a miracle! A little about the miracle: our users visit our site through advertising from different advertising channels, and there are also organic users who went to the site on their own (without spending on advertising) - on the advice of friends, having seen an advertisement in any advertising source, but did not use the link, etc. The cost of attracting users varies from free ("organic users") and almost free - email newsletters, to the most expensive - contextual advertising and advertising on social networks; attracting users through bloggers is in the middle in terms of cost of attraction.

Thus, advertising sources can be arranged in the following order in descending order of cost of attraction:

1) contextual advertising
2) social networks
3) advertising with bloggers
4) email newsletters

Since another company is handling the marketing, it is important for us to conduct our own investigation and we invite you to become our detectives. Your task will be to understand the available sales data and help us make an informed decision regarding marketing campaigns for the next year.

Description of the task:

Link to file .

You work as a product analyst at a foreign company called Delivery Dron, which sells delivery drones online. The company launched sales in four regions in May 2019. Investors are demanding a report on the funds invested, and it is necessary to make plans for the next year.

You need to analyze the available data and answer the customer's main question: in which region and which advertising channel to invest more money to attract customers next year and justify your proposals?

- **Data Description:**

  The file ecom_yl.csv contains data on visits and purchases of users of the website of the online store “Delivery Dron”. Each line represents a user visit. The data is sorted by Session Date.

  Dataset columns:

  0: User Id - user id

  1: Region - region of visit

  2: Device - the device from which the site is accessed

  3: Channel - advertising channel for transition to the site

  4: Session Start - the start time of the user session (login to the site)

  5: Session End - time of end of user session (leaving the site)

  6: SessionDurationSec - session duration in seconds

  7: Session Date - date of visit

  8: Month - month of visit

  9: Day - day of the day of the week of the visit

  10: Hour of day - hour (from 0 to 24) of visit

  11: Order Dt - purchase date

  12: Revenue - purchase amount

  13: Payment Type - payment type

  14: Promo code - use a promo code for a 10% discount (yes-1, no-0).

- **Exploratory data analysis and preprocessing:**

  To conduct the study it is necessary:

  - bring column names in line with PEP8

- ○ check the data for gaps and fill them in if necessary. Justify the decision.
- ○ check data for duplicates, including implicit duplicates in categorical data
- ○ convert the data type of date and time columns to the appropriate format if necessary (pd.to_datetime())
- ○ add a column with the total purchase amount taking into account the application of a promo code for a 10% discount
- ○ determine the period under study, whether to take the entire period for study or not?
- ○ check data for outliers and data adequacy
- ○ add a column indicating the time of day of the visit (morning 06:00-09:59, day 10:00-16:59, evening 17:00-21:59, night 22:00-05:59)
- ○ add a "payer" column with information about whether the user is paying or not.
- ○ Conduct analytical and graphical analysis of data:
  - ■ Share of sales by region
  - ■ Share of sales by source
  - ■ Share of sales by devices
  - ■ Number of users broken down into paying/non-paying by region
  - ■ Number of users broken down into paying/non-paying by device
  - ■ Number of users broken down into paying/non-paying by source
  - ■ Graphs showing whether there is seasonality in sales by month, day of the week, time of day
  - ■ Chart of the number of purchases by payment type

● Carrying out calculations

- ○ Calculate the average bill.
- ○ How many purchases does 1 user make on average?
- ○ Calculate the average session duration by advertising channels.
- ○ Calculate average session duration by device type.
- ○ Determine the top 3 advertising channels by average check.
- ○ Determine the top 3 regions by average check.
- ○ Determine the top 3 months by average check, broken down by region.
- ○ Calculate MAU* for each month, broken down by advertising channels, and identify the top 3 advertising channels by the number of unique users per month.

* MAU (Monthly Active Users) is the number of unique users per month.

- ○ Create a table that will indicate by advertising channels: number of users, number of unique users,

the number of paying users, the amount of sales and determine which source “brought” the most paying users and the largest amount of sales.

- Hypothesis testing
  - Does device type affect the number of purchases per day in each region?
  - Does the type of advertising channel affect the number of purchases per day for each region?
  - Test the hypothesis that the average bill differs depending on the region?
  - Test the hypothesis that the average bill differs depending on the advertising channel?
  - Test the hypothesis that the average bill differs depending on the time of day?
  - Is there a relationship between session duration and purchase amount?
  - Come up with and test at least 2 more hypotheses based on the team’s hypotheses.

- Regression modeling
  - Build a regression model of product sales based on sales factors. The choice and argumentation of factors remains with the team.

- Dashboard
  - Create a dashboard that includes the main metrics of the online store: indicators, graphs and diagrams (at least 5), which will answer the following questions from investors:
  - sales amount, average check, number of users, number of paying users, average session duration
  - distribution of users by regions and devices
  - graph(s) of the number of users and the amount of sales by month
  - Key diagrams - the total sales by advertising channel and the number of attracted paying users broken down by region.
  - Add a filter to the dashboard by advertising channel, date, region type and payment type.

- Conclusions
    - It is necessary to present key findings and insights in a presentation with visual representation (up to 10 slides, not counting the title and final slides).