**AI 620 Emerging Topics in Artificial Intelligence**

**HOS08A Predict Ad Click Conversion through XGBoost**

03/28/2023 Developed by Yared Shewarade

09/23/2024 Reviewed by Anh Nguyen

11/19/2024 Reviewed by Jonathan Koerber

School of Technology and Computing (STC) @City University of Seattle (CityU)

**Before You Start**

* The directory path shown in screenshots may be different from yours.
* Some steps are not explained in the tutorial**.** If you are not sure what to do:
  1. Consult the resources listed below.
  2. If you cannot solve the problem after a few tries, ask the course student worker for help.

**Learning Outcomes**

Students will be able to learn:

* Introduction to Ad Click Conversion through XGBoost
* Practice on Mobile Ads Click-Through Rate (CTR)

**Resources**

* Tripuraneni, S., & Song, C. (2019). *Hands-on artificial intelligence on amazon web services: Decrease the time to market for AI and ML applications with the power of AWS* (1st ed.). Packt.

# Introduction to Ad Click-Conversion through XGBoost

**XGBoost** is a gradient-boosting library that is very popular in the ML community. It's a great library to start with and is very easy to use. It's also very fast and can be used for classification and regression problems. The Internet advertising conversion rate is an important quantitative indicator for search engine service providers and advertisers, the realization of Internet advertising conversion rate prediction under the big data platform has strong theoretical research value and practical application value. Since the conversion of internet advertising is a small probability event under a large amount of data, therefore, to increase the advertising conversion rate prediction, a XGBoost-based method is proposed. Through the analysis of large-scale advertising conversion logs, we extracted data features and constructed data sets, then applied XGBoost algorithm to achieve advertising conversion rate prediction successfully. The experimental results show that compared with the traditional machine learning methods in industry, XGBoost has better prediction results than other methods under the same features extraction and data sets.

# Practice on Mobile Ads Click-Through Rate (CTR)

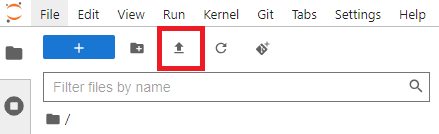
**Note:**

* Open the notebook using SageMaker.
* For submission, take the screenshot for all steps and save it in your local repository along with your code.

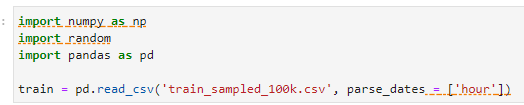
1. **Type the following Import the required module and upload the data.**

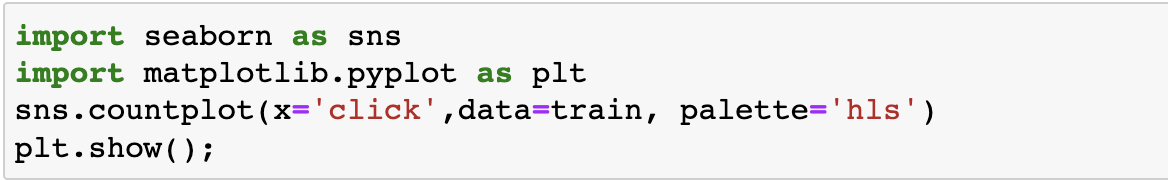
The original dataset is from a Kaggle competition ([here)](https://www.kaggle.com/c/avazu-ctr-prediction/data) with a train dataset of more than 4 million records. The provided “train\_sampled\_100k.csv” file from the HOS10A repository is a reduced-size file that was sampled from the original dataset.

Upload the “train\_sampled\_100k.csv” to your notebook.



The dataset contains over 40 millions of records, but we will use 100 thousands records only.



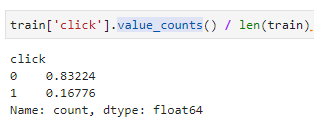
****

**Chart, bar chart

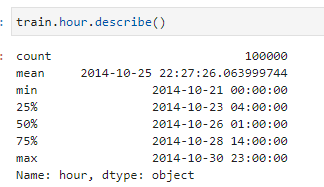
Description automatically generated**

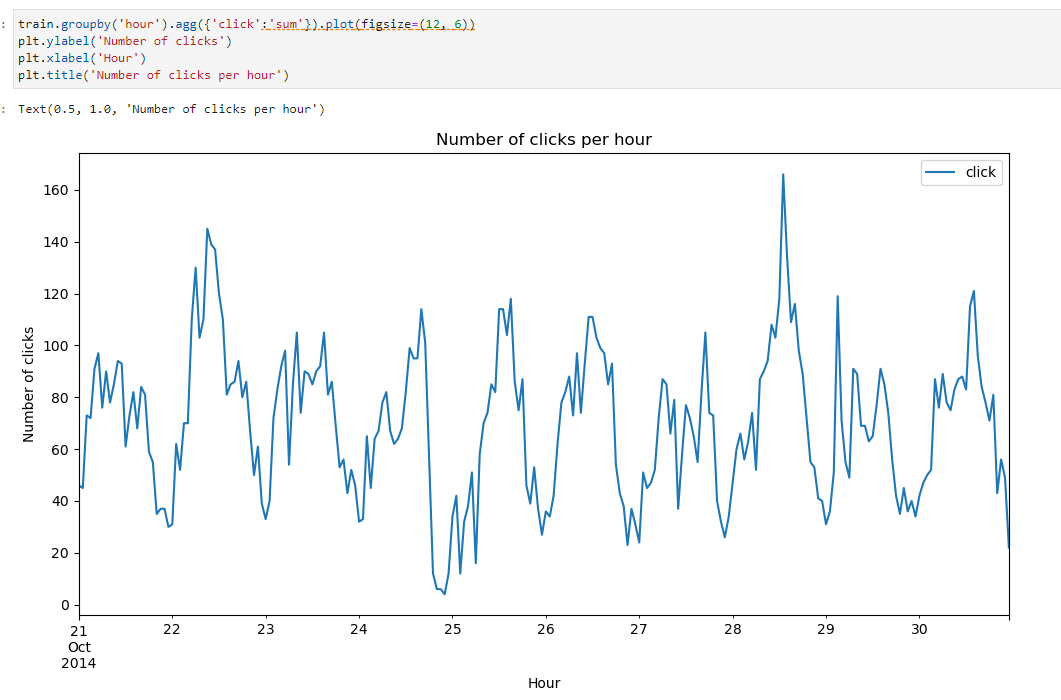
1. **Type the following to apply the Ads Click**

Let’s see the click status.

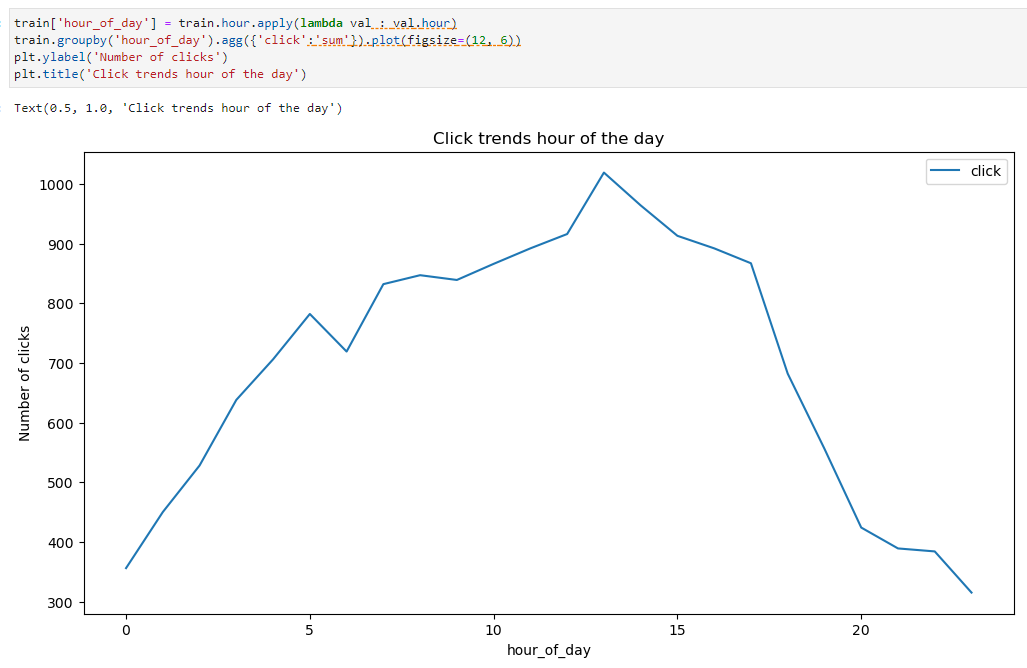


Note: - The overall click through rate is approx. 17%, and approx. 83% is not clicked.



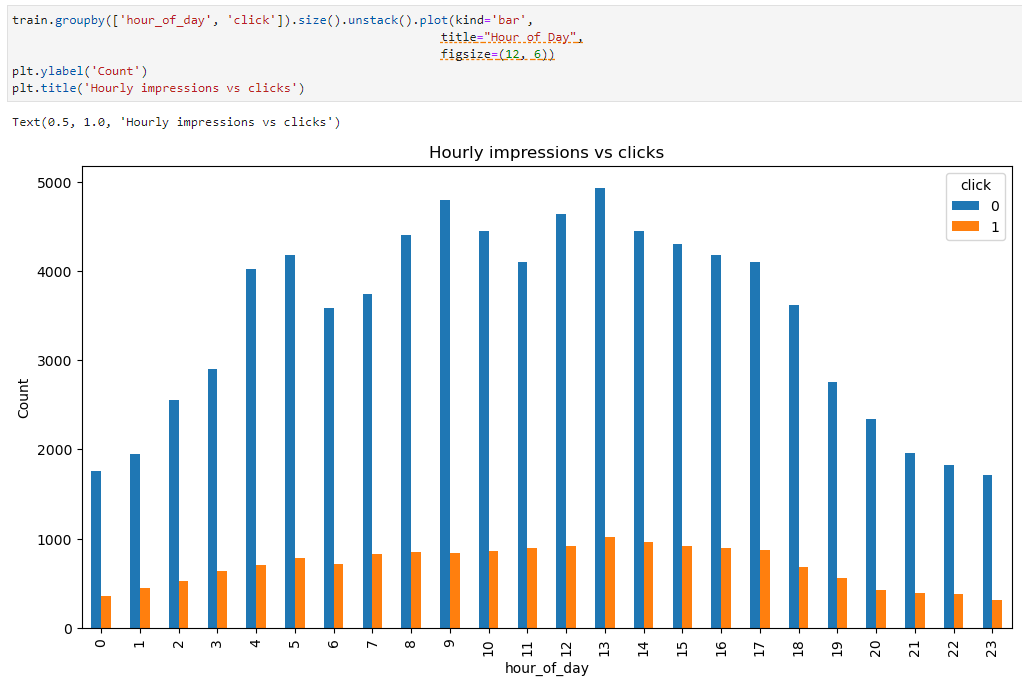


The hourly clicks pattern looks pretty similar every day. However, there were a couple of peak hours, one is sometime in the mid of the day on Oct 22, and another is sometime in the mid of the day on Oct 28. And one very low click hour is close to mid-night on Oct 25.



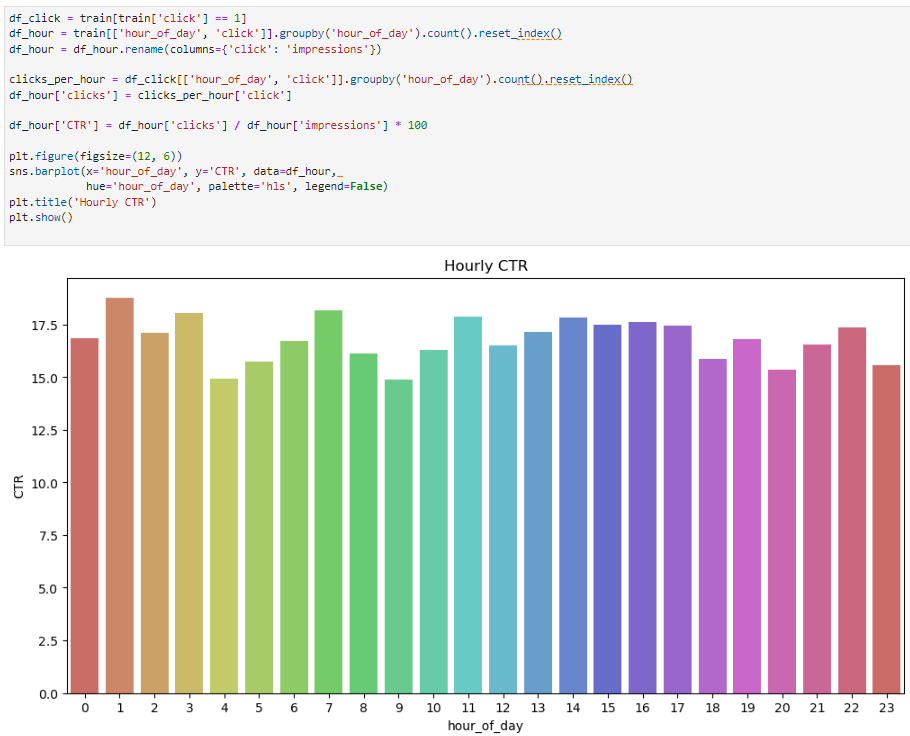
In general, the highest number of clicks is at hour 13 and 14 (1pm and 2pm), and the lowest number of clicks is at hour 0 (midnight). It seems a useful feature for rough estimation.

Let’s take impressions into consideration.



Now that we have looked at clicks and impressions. We can calculate click-through rate (CTR). CTR is the ratio of ad clicks to impressions. It measures the rate of clicks on each ad.

**Hourly CTR**



**HOS submission instructions:**

1. Please install the GitHub Desktop: <https://cityuseattle.github.io/docs/git/github_desktop/>

2. Clone, organize, and submit your work through GitHub Desktop: <https://cityuseattle.github.io/docs/hoporhos>