**Attribute Selection**

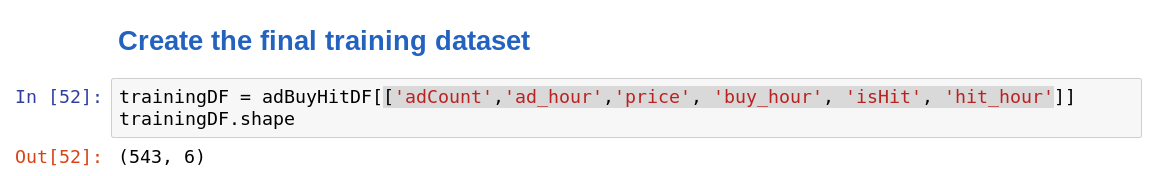
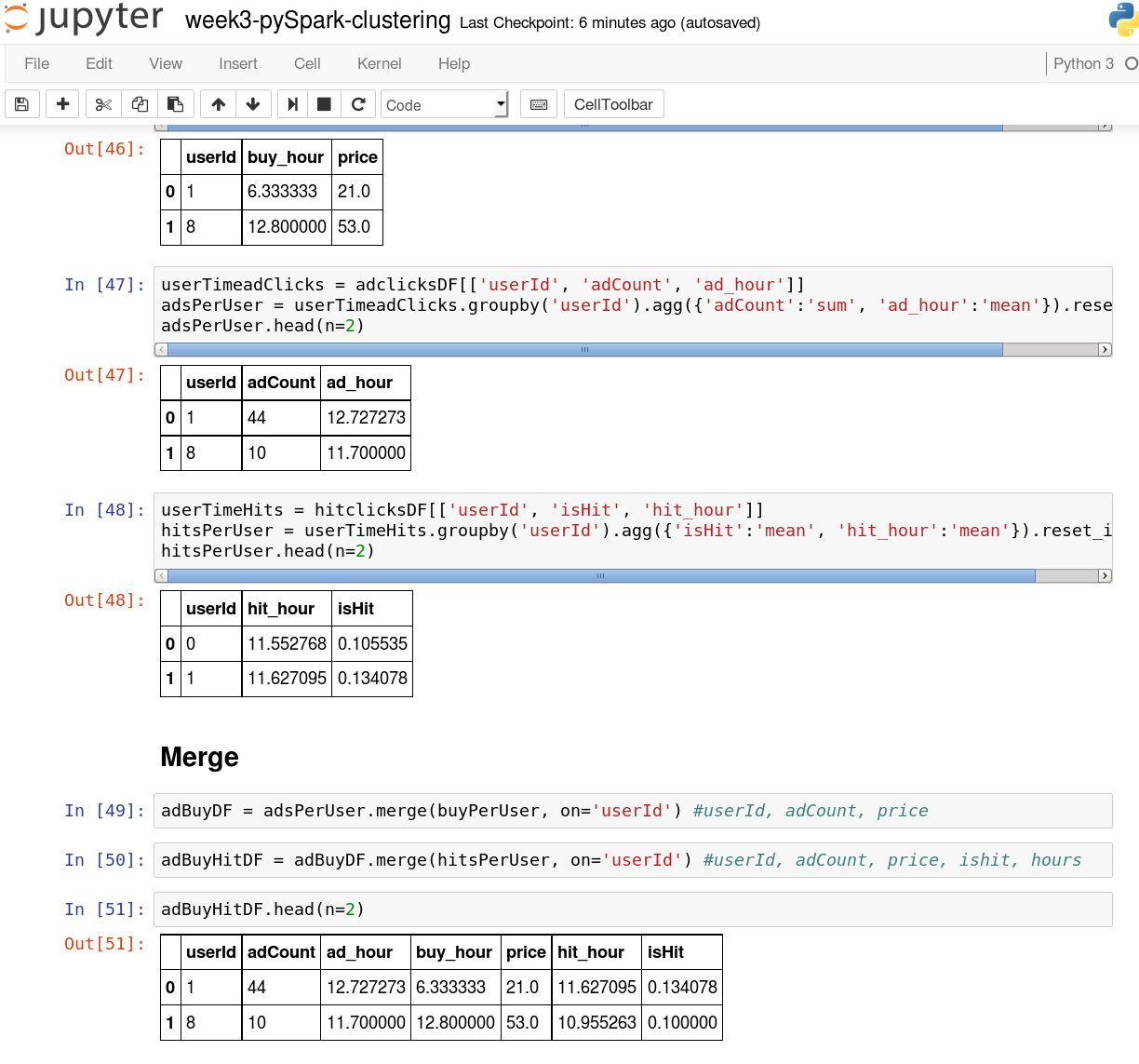
features\_used = ["timestamp", ”price”, “isHit”, “adCount”]

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Selection** |
| "buy\_hour"  "ad\_hour"  "hit\_hour" | Users are more likely to purchase at a certain time period like evening or afternoon. We extract hour from time stamp of buy, ad and game hit |
| “price” | Buy: Price is already an important factor which correlates with buyId |
| “adCount” | Ad: count of ads |
| “isHit” | Game: Compute average hit per user id before merging to know skill |

**Training Data Set Creation**

The training data set used for this analysis is shown below (first 5 lines):

**Below is the detailed merge operation to create the training data frame.**

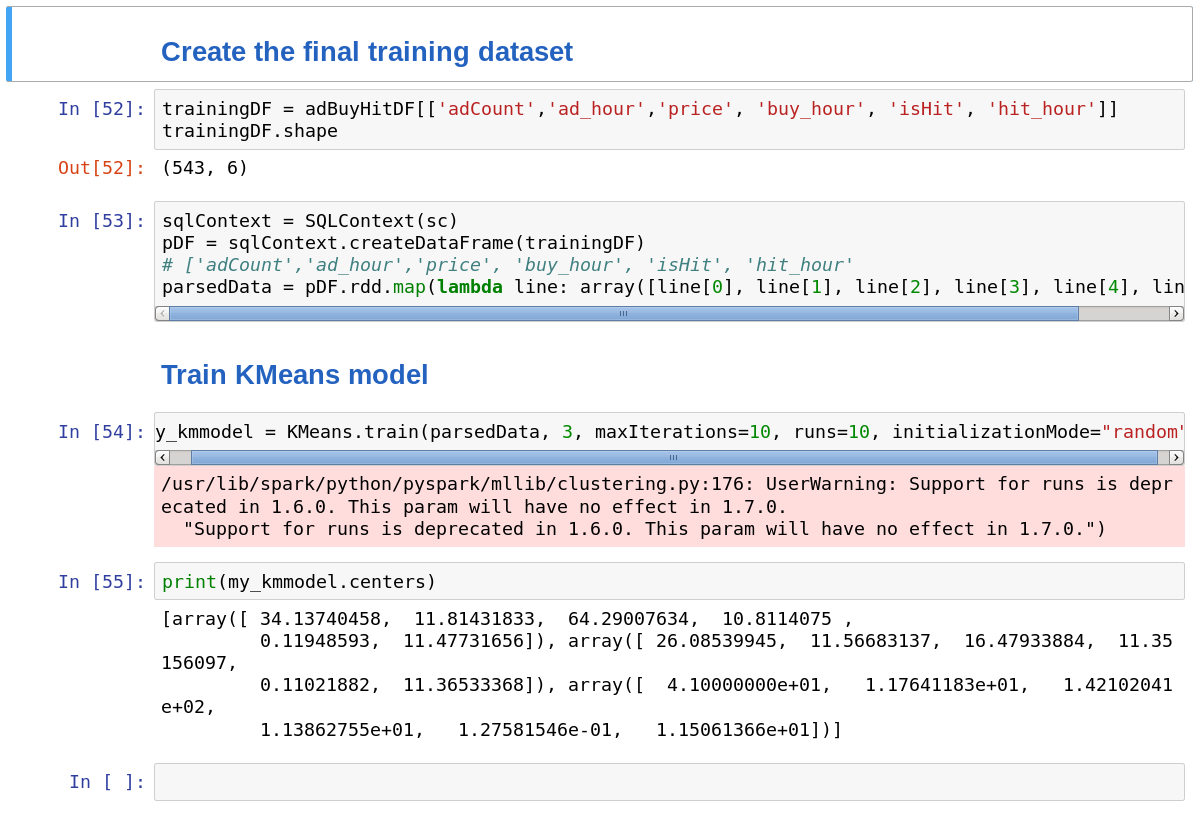


Dimensions of the final data set:  **(543, 6)**

# of clusters created: **k=3**

**Cluster Centers**

The code used in creating cluster centers is given below:



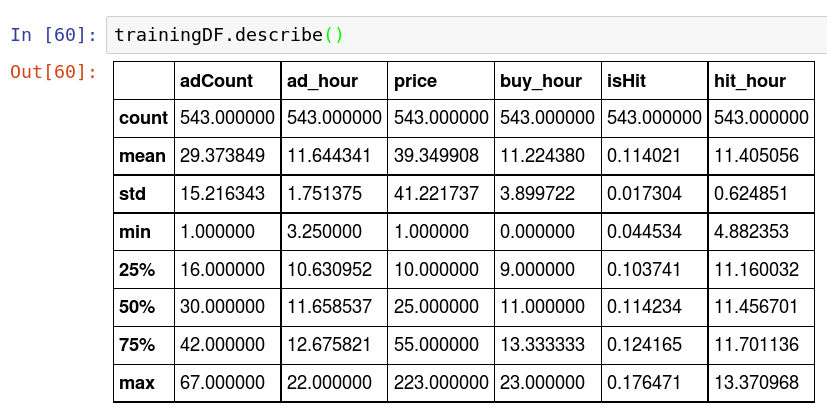
Cluster centers formed are given in the table below

|  |  |
| --- | --- |
| **Cluster #** | **Center ['adCount','ad\_hour','price', 'buy\_hour', 'isHit', 'hit\_hour']** |
| 1 | [ 34.13740458, 11.81431833, 64.29007634, 10.8114075 ,  0.11948593, 11.47731656] |
| 2 | [ 26.08539945, 11.56683137, 16.47933884, 11.35156097,  0.11021882, 11.36533368] |
| 3 | [ 41.0000000 11.7641183, 142.102041, 11.3862755,  0.127581546, 11.5061366] |

These clusters can be differentiated from each other as follows:

* Cluster 1 is different from the others in that users with intermediate result in game clicks, adCount and price are grouped at this cluster center.
* Cluster 2 is different from the others in that the users with minimum adCount and average price also have minimum average hit rate
* Cluster 3 is different from the others in that the users with maximum adCount also have the highest hit and average price leading to higher revenue.

Below you can see the summary of the train data set:



**Recommended Actions**

|  |  |
| --- | --- |
| **Action Recommended** | **Rationale for the action** |
| Propose ads right after a winning streak, level up or good hit rate | Hour extracted from time for ad\_hour>hit\_hour>buy\_hour as observed in the cluster centers |
| Select the high hit and revenue generating customer and show them ads later when compared to others | High ad count relates to high price and hit rate |
| Selectively offer training and for low hit rate players so that they can engage and generate more revenue | Low hit rate clusters with low revenue and low ad count |