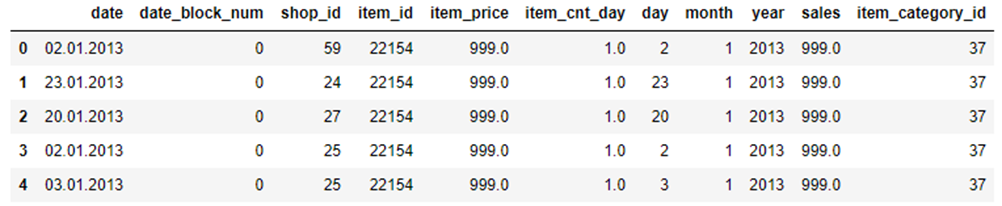
# **Group 5 - Assignment Summary**

## **Abstract**

Objective of assignment report is to build a prediction model and predict the total sales for every product and store in the next month for the 1C Company. The dataset chosen on which the analysis is "Predict Future Sales" dataset. The dataset was downloaded from the public dataset on Kaggle at the URL, <https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data>.

Key dataset includes **“Sales\_Train\_V2”** which provides item wised month-wise sales information of the shop. It also includes price and the item category. Dataset contain close to 3 million records. Other dataset includes, shops.csv, item\_categories.csv, items.csv, test.csv.

**Data Analysis -** Initial analysis revealed the little need for the data cleansing. Shop name, item name and all the text within the dataset are in the Russian language. However, it doesn’t limit implementing and executing the prediction model. Create new data elements such **Year, Month and Day** fields to help time series analysis and **sales** elements from item\_price and item\_cnt\_day. Merged the items\_category field from item\_categories to manipulate category wise sales.



**Feature Analysis**

**Seasonality Summary**

* **Sales by year** - Year trend chart shows that 2014 total sales are higher. 2015 only contains 10-month data, so it is not comparable with the other 2 years.
* **Month wise Sales - Analysis** of e Bar Chart & the Boxplot show that the Total Sales in the month of November and December are higher compared to other months.
* **Sales by Day -** Both the Bar Chart and the Boxplot show that the day 19 and day 30 have highest total sales.

**Sales and volume by shops**

* The Bar Chart shows that Shop ID 31, 25, 42 has the top 3 total sales.
* The box plot suggests that Shop ID 9 and Shop ID 20 has the highest daily sales for single item
* The Bar Chart for volume shows that Shop ID 31, 25, 54 has the top 3 total sales volume.
* **Sales by Day -** Both the Bar Chart and the Boxplot show that the day 19 and day 30 have highest total sales.

**Model used and execution**

**Observation**

**Conclusion**