# Project Type: Data Analysis & Visualization Project

# Project Name: Electronic Products Catalog

# Document Type: Risk & Issues Analysis

# Group 7

Astha Rastogi

Ankita Roy

Eric Pozholiparambil

Shweta Pathak

# Project Information:

Our Project deals with about 15000 electronic products with their respective brands, availability, sales and other characteristics such as whether the product is on sale and involves shipping cost or not.

We encountered certain issues while we were executing the project. They are all mentioned below along with the measures that we took to handle them and successfully execute the project.

# Issues Addressed:

Choice of No-SQL Database:  
We had a lot of choices to select from the given No-SQL database. However, we wanted to choose the database based on 2 criteria’s:

1. Characteristics of No-SQL DB (Scalability, Ease of use etc)
2. Connectivity Options for Microsoft Power BI

After careful consideration, we decided to go ahead with Mongo DB because we found good amount of professional support online, elaborate and concise documentation and our DBA member in the team had good knowledge of the same.

## Understanding Given Data:

The Data that is given to us has less than 16000 rows, out of which post cleaning, only about 14000 rows remain. The second issue we encountered was the interpretation of the last 3 unnamed columns. However, the good part was after analyzing this with Python, we realized that all the data in those columns was null and hence we removed it. The next problem was the interpretation of some of the columns such as ‘ID’ and ‘Keys’.   
Analyzing and cleaning the data helped us understand which column actually represents what data.

Database Connectivity to Power BI:  
This issue persisted for a very long duration throughout the project. We figured there were about 3 ways to connect Mongo DB:

1. Using REST API Code
2. ODBC Connector
3. Using R script

However, after careful apprehension of the above 3 ways, we realized ODBC Connector was the most preferable and easy way. Once the ODBC Connector was configured (which took the most time), we connected the data to Power BI.  
3 important lessons we learned here:

* Configuring the correct port number for the connection
* Testing the database connectivity through the ODBC Connector first (for which data needs to be loaded in the database prior)
* Which Commands should be used alongside Mongo DB for the ODBC Configuration.

ODBC Connector Choice:  
There were Several ODBC Connectors available such as Simba and DataPoint, we chose the one which was provided by Mongo DB on their official Site. This choice was unanimous because the documentation and steps were clearly provided on their site and the git hub link.

* Documentation: <https://docs.mongodb.com/bi-connector/master/reference/odbc-driver/>
* Link for download: <https://github.com/mongodb/mongo-odbc-driver/releases/>

Visualization of the given data:  
With the lack of any geographical or spatial data as well as time data, we had no access to any trend lines or time analysis for our visualization. Thus, for visualization we had to keenly observe the data, note the important features and factors affecting one another and then derive an inference from it.  
Our data mostly provides us with information about products and their sales which can be characterized due to different factors such as its brand, its availability, how new the product is and what type does it belong to.

---