# 

# Industry Project Proposal

|  |  |
| --- | --- |
| Name of the Organization | GroupBy |
| Project location (city) | Austin, TX |

**GroupBy** is an ecommerce start up that aims to create the world’s most relevant and highly converting site experience, maximizing revenue through e-Commerce channels all while requiring less effort and cost to build and run.

### Project Description

**Problem definition**

*[50-100 word description of the problem which the candidates need to solve]*

|  |
| --- |
| With the surge in online shopping brought about by the recent pandemic, the requirements of online shopping platforms to be both user friendly and accurate for predictive inventory and stocking applications has become significant. In this project, the primary objective is improving the predictive effectiveness of the online shopping experience for *essential products* that range from *vitamins, supplements, OTC medicines to home cleaning supplies*.  Our business unit has a hypothesis that if the probability of the customer conversion is less than 0.2 it doesn’t matter what we do, they will not buy and if it’s more than 0.8, we don’t want to change a lot in their journey. Our intention is to **personalize the shopping experience** for customers for whom the conversion probability lies in the range [0.2-0.8]. With several freely available E-commerce data sets online such as:  1) [Kaggle E-Commerce events history in cosmetics shop data set](https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop)  2) [Online Retail Data set](https://archive.ics.uci.edu/ml/datasets/online+retail)  The business unit is interested in optimizing the customer’s journey in one of its retail customers. More specifically the business unit wants to know **Are we able to identify conversions for the users?** |

**Key Research Questions/ Technological constraints that the Project will Answer**

|  |
| --- |
| 1. Is it possible to predict customer conversion (from viewing to purchasing)? 2. Can we personalize the experience from customer-product interactions? 3. Does the product catalogue information help in improving customer experience? 4. Can personalized shopping experience be predictively modeled? 5. Can we apply transfer learning from other E-commerce data sets to our local data set? 6. Can we identify “new customers”? If yes then what is the best strategy to work with them? 7. Is there a pattern and trend across product types regarding the conversion event? 8. What are the top features that define customer conversion? 9. Can we explain and visualize the predictive model for customer conversion on a journey and at a session level? |

**Final deliverables at the end of the project**

*[Please list the desired technical deliverables from the project team in as much detail as possible]*

|  |
| --- |
| 1. A predictive model for viewing and purchasing customer conversion analysis. 2. A recommendation system for product personalization 3. Complete analytics and explainability of results and outliers. 4. Technical reporting to the business unit |

**Key activities/ technologies the project team may be expected to undertake/ work with**

*[E.g. What kind of technology stack they will work with, the datasets they may need to work on, what kind of analysis they may be expected to undertake, etc.]*

|  |
| --- |
| 1. Data: 2. Start with several freely available E-commerce data sets online such as:   1) [Kaggle E-Commerce events history in cosmetics shop data set](https://www.kaggle.com/mkechinov/ecommerce-events-history-in-cosmetics-shop)  2) [Online Retail Data set](https://archive.ics.uci.edu/ml/datasets/online+retail)   1. Work on in-house anonymized data from Groupby. Our intention is to leverage the learnings from such E-commerce data streams and assess transfer learning capabilities on our custom inhouse data set. A snapshot of the data schema is provided in Appendix A.   2. Tensorflow, Flask, ONNX |

**Expected learning outcomes**

*[What do you expect the candidates to learn from the project. Please mention the technical skills they will imbibe over the project.]*

|  |
| --- |
| 1. ML model development 2. Benchmarking, transfer learning 3. ML model deployment 4. ML model optimization |

|  |  |
| --- | --- |
| Desired Team Size (if any): | 3 |