RAZOR GROUP

Case Study – Data Engineers

January 2022

TECHNOLOGY & DATA ANALYTICS

Razor Group is hiring for various positions in Data Engineering, Data Science and Business Intelligence!

Help us leverage data to pursue novel approaches to acquire and grow strong brands that consumers love. Unlock data silos along the E-Commerce value chain to drive truly data-driven impact on one of the world's largest industries

We are looking for amazing personalities in the field of Data Science, Data Analytics or Business intelligence, ideally with experience in E-Commerce. We are looking for both young and ambitious, as well as well-experienced experts. At Razor, you will be able to prosper as an individual contributor or as a leadership personality

RAZOR GROUP

“We are on a mission to build a new-age consumer goods giant. Starting by acquiring and scaling highly profitable Amazon FBA assets on global scale..”

You can read more about us in this article from TechCrunch. Click here!

CASE STUDY

Build data model for analyzing revenue & fees of Amazon brands

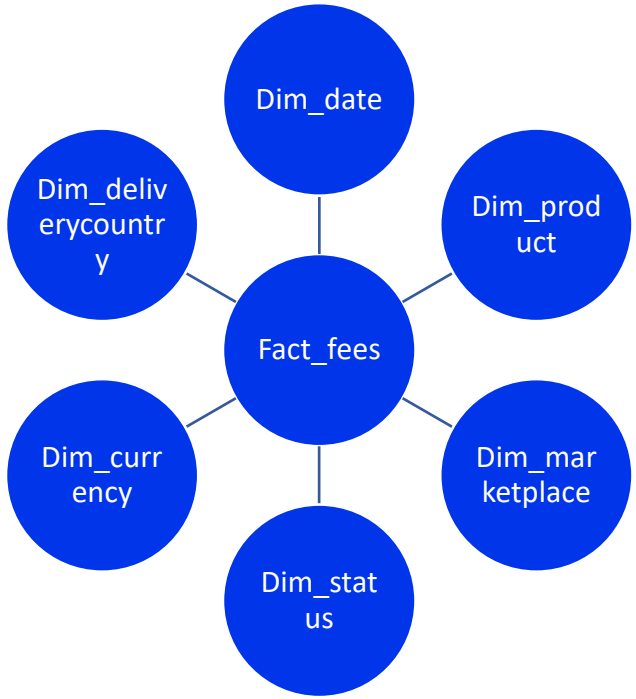
CONTEXT	*As you might know already, we **own & operate 70+ Amazon brands** and it is very important for our brand managers to be aware of their brands performance so that operational decisions can be taken real-time. As with any data source, the raw data received from Amazon contains a lot of data noise and not easy to consume. As part of this task, you may need to **parse and sanitize these raw datasets** (orders and fees data) and **develop a relational data model with fact and dimension tables** to enable the Razor Brand Managers to consume this data in an efficient manner*
PROBLEM STATEMENT	*This assignment comes with **2 raw datasets of Orders and Amazon fees** on the next slide. Your task is to develop an **automated ETL process** for sanitizing, parsing and integrating this data in a Data Warehouse (on any stack of your choice) and build a relational data model as mentioned in the schema diagram shown in the next slide as per the Kimball principles. Note that SCD (Slowly Changing Dimensions) type-2 dimensions should be implemented for fact_orders using "change data capture" method*
ASSESSMENT	*Note that the above datasets are **randomized sample datasets** with no business impact, and this task is meant to assess your ability to work with raw datasets and your approach to solve this* *We will assess you on the **data flow design process** and **comprehensiveness of the ETL jobs** (SQL queries or python/spark scripts), and **compliance to Kimball principles** in developing the data model. We will ask you to present the data models along with the E/R diagrams in a joint call via screenshare*

RAZOR GROUP

Data sets for the task

This assignment comes with 2 datasets, Orders-Oct'21 with raw orders data for the month of October and incremental orders with orders data **from 10/25 to 11/7**. Please develop your **ETL process to handle incremental data loads daily with last 2 weeks data**. Please find below the recommended fact and dimension tables, but please feel free to modify them as needed to best suite the data model. Please make sure the surrogate keys are non-business keys and should be native to the Data Warehouse

Download the data sets Orders: Oct'21 Orders: 10/25-11/7 Fees: Oct'21 Fees: 10/25-11/7

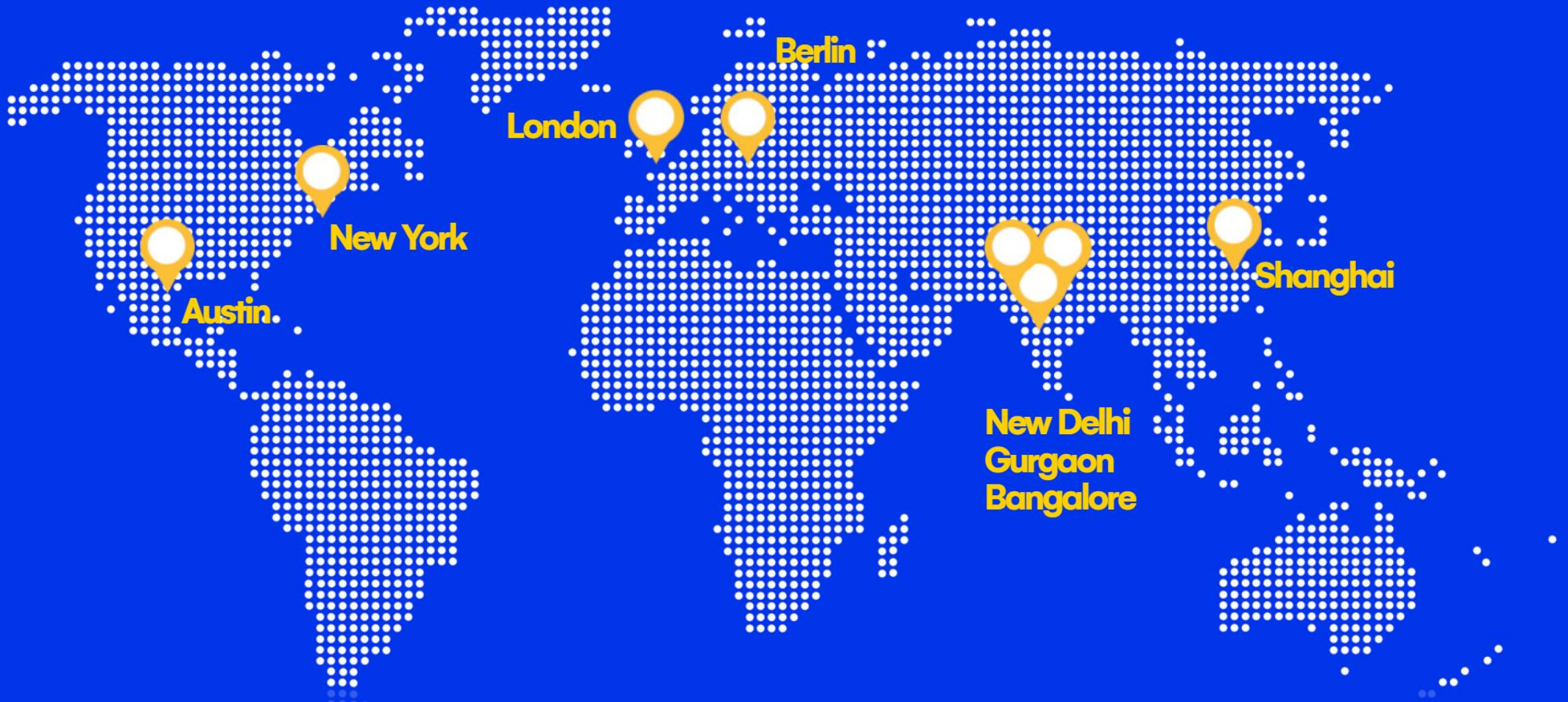


RAZOR GROUP

Guidelines

- Raw data in the orders dataset is at order-item level hence each order can have multiple rows
  - ‘sku’ is the unique item identifier
  - ‘asin’ is the product level identifier
- ’sku’ field is not clean in the raw data, and you will need to clean up this data before joining with other tables and feeding the data model
- Note that each order in the settlements data may have multiple fee related line items and you will need to represent the overall fees charged for each order with granular level fee item information
- Data needs to be cleaned (remove duplicates, remove special characters, format to be consistent, enrich to be intuitive, etc.) before pushing them to the fact and dimension tables
- Field names has to be intuitive for business users in the fact and dimension tables
- Data models should allow the users to query the atomic level data
- Please create a dim_date data and use date keys instead of date/datetime in the fact tables
- Please make sure fact and dimension tables are build as per Kimball principles

RAZOR GROUP



08/2020	300+	$ 550 M+	6	70+
FOUNDED	EMPLOYEES	FUNDING	LOCATIONS	BRANDS