Avi Skidelsky, DAV-6100, M5 Assignment

1. The grain of the data here is the individual transaction for each product. We don’t know exactly what types of analysis we will do in the future but it will likely have to do with sales so the actual transaction, i.e. the sale, captures all of that data required.
2. The dimensions associated with each transaction are the date of the transaction, product being sold, store that the transaction took place in, the promotion that may have been applied to this transaction, the cashier who processed the transaction, and the payment method used to complete the transaction. The facts for each transaction are the quantity of the product, the regular price of the item, the possibly discounted price of the item, the net unit price of the item, the extended discount dollar amount of the item, the extended sales dollar amount of the item, the extended cost dollar amount of the item, and the extended gross profit dollar amount of the item. All of these help paint the picture of the aggregated numbers associated with each transaction.
3. Since the dimensions aren't naturally correlated with each other the many to many relationships between the keys can get quite complicated seeing as every store is open every day and sooner or later almost every product will be included in a promotion. The person modeling this data, in an attempt to normalize everything without moving into a snowflake schema, might end up creating one giant normalized fact table that is hard to read and analyze as opposed to smaller, more concise, and more easily queried fact tables.

A picture containing diagram

Description automatically generated