ANA 655 - Database Design for eCommerce Health and Wellness Business

**The Wellness Team**

Students: Aditi Bhujbal, Stefan Francisci, Shila KC, and Jennifer Knight

Evaluator: Barbara Lauridsen,

July 1 earned 47 see page footnote for “Payment\_Dim”.

Project - Week 4 - Part 3

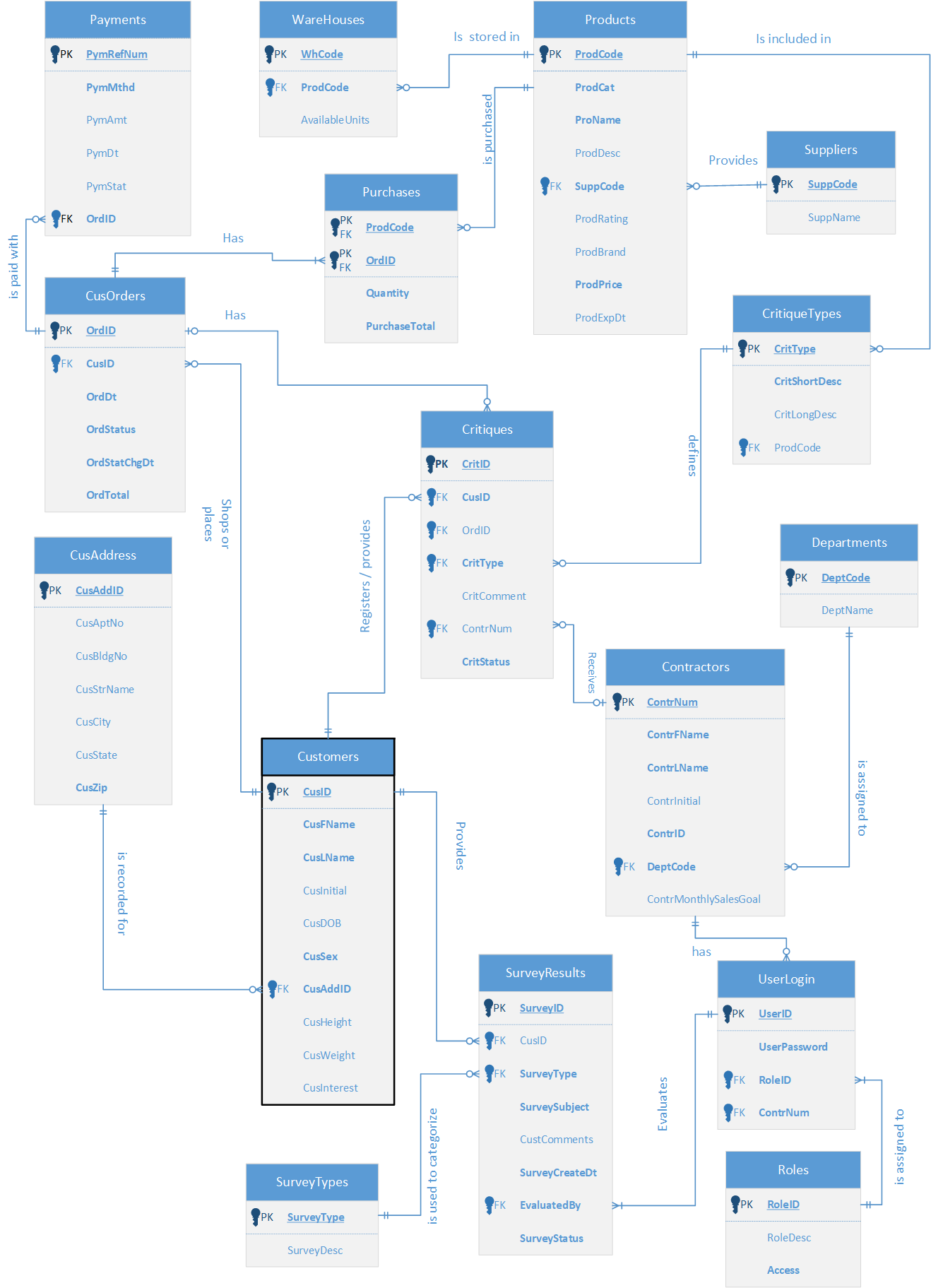
June 30th, 2022 – version 2

**REFINED OPERATIONAL ERD:**

The following diagram shows The Wellness Team’s most recent Operational ERD, which has been refined to fulfill Data warehouse requirements.

Modifications made to previous model:

1. Modified name of all entities to be plural e.g., Products, Employees, CusOrders.
2. Added new attributes in Purchases table – Quantity and PurchaseTotal.
3. Added new attributes to Products table – ProdBrand and ProdPrice
4. Added new attribute to CusOrders table – OrderTotal.
5. Removed ‘PymMthd’ attribute from CusOrders table and added ‘OrdID’ to Payments table to connect the two tables.
6. Added new attribute to Payments table - PymRefNum and set it as primary key.
7. Removed ‘primary key’ constraint from ‘PymMthd’ attribute in Payments table



*A larger version is included in the appendix.*

**WAREHOUSE SCHEMA(STAR):**

The image below shows the star schema design of a data warehouse prototype that we proposed to design in this project. *A larger version is included in the appendix.*

Graphical user interface

Description automatically generated

The scheme was created by The Wellness Team. We referred to a paper by M. **Drozdowicz** provided by instructor Dr. Barbara Lauridsen.

This star schema consists of **4 Fact** and **6 Dimension** tables.

**Fact tables:**

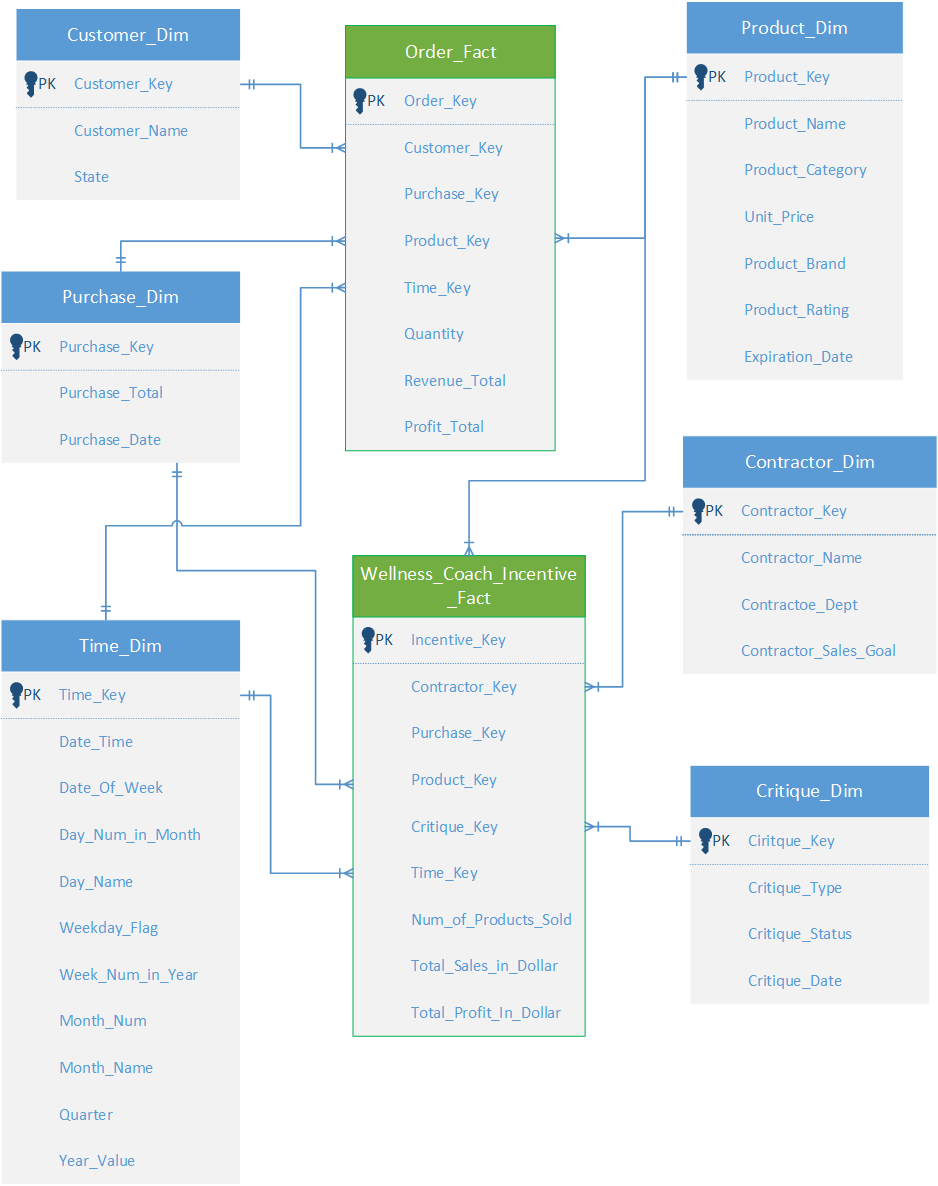
1. **Survey\_Fact**: This table will store Survey\_ID as PK and the measures of customer surveys along with their response (Positive/negative) to recommendations provided by wellness coach. **Customer\_Key, Time\_Key, Contractor\_Key** are numeric keys that will connect this fact table to their respective dimensions. **Received\_date** will be populated from SurveyCreateDt in SurveyResults of operational database. **Recommendation\_sent** (Yes/No) and **Customer\_Response\_Ind** (positive/negative) will store the transformed data from SurveyStatus attribute of the same entity in operational db. We are undecided as to whether we will keep this Fact table in our final design or not. We have chosen to include it here as a possibility, but may remove it after prototype testing.
2. **Order\_Fact**: Data from CusOrders table of operational database will be extracted into staging after orders achieve a status of completed (either delivered or abandoned), where they will be transformed and loaded into this table to provide measure metrics such as “Total Number of orders in each status (delivered/returned) per day”, “Top 5 products bought by customers per quarter”, “Product(s) that have contributed to maximum returns in a month”, “Total revenue and profit of a year/quarter/month. Numeric keys will be stored to describe customers, product and payment details associated with the order, and Time\_Key will contain information about the date and time of creation of a record into Order\_Fact.
3. **Critique\_Fact**: This table stores the data extracted from the Critiques table of our operational database. The data stored in Critique\_Fact table will be useful to determine number of completed critiques (such as complaints or compliments), number of endorsements received for a product or services provided by a wellness coach (who is a contractor). Critique\_Fatc\_Key is a primary key to distinguish between each record in this table. Numeric key attributes are foreign keys to connect to primary keys of Customer, Product, Contractor, and Time dimensions for the purpose of searching history for patterns of activity seeking possible resolutions. This information will be loaded into the data mart on a monthly or quarterly basis after the critique is marked completed in the operational database. This information will be analyzed to view customer trends and then evaluated to determine if there are any problems (such a defective product or unprofessional employee) the business needs to address on a global level.
4. **Inventory\_Fact**: The source of the data stored in this fact table is **Warehouses** table from operational database. It provides an insight into the quantity of each product on hand for each Warehouse when transactions were loaded into the data mart, Product\_Key, Supplier\_Key and Time\_Key attributes connect the respective dimension tables to describe each item stored in this fact table. This data plays an important role in inventory management decisions. For example, if inventory historical transactions are migrated on a periodic basis, such as weekly, a metric quantity-on-hand could be illustrated as product turnovers, depletions, or supplier’s involvement when a shipment arrives and is scanned into a Warehouse. The primary key may become the way for a user to locate the original Inventory records in the source database if it decided to assess process improvement opportunities.

**Dimension Tables:**

1. **Customer\_Dim:** This tablewill containonly the required information for pulling queries about completed transactions generated by customers such as Customer Name, Age, Sex from Customers table and City, State, Zip from the CusAddress table in operational source. We will not track individual address because that level of information is unnecessary at this level, but will include more general customer address information so we can query sales by geological region. This information will be especially helpful when we run marketing campaigns. Customer\_Key is a primary key to uniquely identify each customer, usually to track purchases made and may be a match to Customer Key in Ops source for completed transactions which are loaded as facts. This data can be used to identify number of customers in each age group (young, adult, and old), number of Male and Female customers, grouping customers according to the state they reside in. This data will also account for analyzing measures in fact tables it is connected to.
2. **Contractor\_Dim:** The source of the data in this table is Contractors from our operational source. Primary key Contractor\_Key is used as foreign key in Survey\_Fact and Critique\_Fact to describe contractors associated with the measures in them. It also contains Contractor\_Name and the Department they belong to. This data can also be used to analyze the performance of contractors assigned in each department.
3. **Payment\_Dim:** <Purchase\_Dim?? see comments in footnote [[1]](#footnote-2) > This table provides more information about the payment and method used to pay the amount for each order. Payment\_Key is a primary key of this dimension and is also stored as foreign key in Order\_Fact. Attributes Payment\_Method, Amount\_paid, Payment\_date and Payment\_Status <[[2]](#footnote-3)> contains the data extracted from Payments table in Operational source. This data can also be analyzed to identify the total amount received by each payment method or we can identify the method that customers prefer to use to pay for their orders.
4. **Product\_Dim:** This table stores information that is relevant to products such as product name, category, price of a product per unit, brand, aggregated rating, and date of expiration. The Primary key of this dimension is Product\_Key which connects it to Order\_Fact and Inventory\_Fact.
5. **Supplier\_Dim:** This table containstwo attributesat this moment, Supplier\_Key and Supplier\_Name. **Supplier\_Key** is the primary key and connects this dimension to Inventory\_Fact. A question may arise that why a separate dimension just to store Supplier\_Name instead of including the Supplier\_Name in Inventory\_Fact itself. The purpose of doing this is to accommodate future requirements to store location information of suppliers, average rating of suppliers based on rating to a product they supplied. This new data will be used to map the supplier to each product after identifying the supplier who provides the best quality of a product.
6. **Time\_Dim:** Time dimension is important in a data mart to identify the most recent data loaded to it and is also useful in retention of data. In our model, Time\_Dim contains Time\_Key which is a primary key and will be recorded in each fact table with the most recent value. It also contains attributes to describe the time measurement such as Date and Time, day of a week, day number in a month, name of the day in a week, whether it is a weekday or holiday, Month, Quarter, Year, etc.

**SALES INCENTIVES SCHEMA:**

The image below displays a star schema that would be used to derive key performance indicators regarding wellness coach performance. From this data mart, we can view total recommendations and complaints related to specific wellness coaches as well as their sales performance. Conceptually, this was created to provide a monetary bonus for wellness coaches to earn by providing strong customer service and attempting to reach a monthly sales goal.



*A larger version is included in the appendix for closer viewing.*

This is a conceptual Star Schema from the Wellness Team that in practice could be used to determine agents who had no complaints and achieved approximately 150 sales per quarter. They would thus be considered top performers and would qualify for incentives as seen in the Wellness Dashboard further below. The breakdown of the tables are as follows:

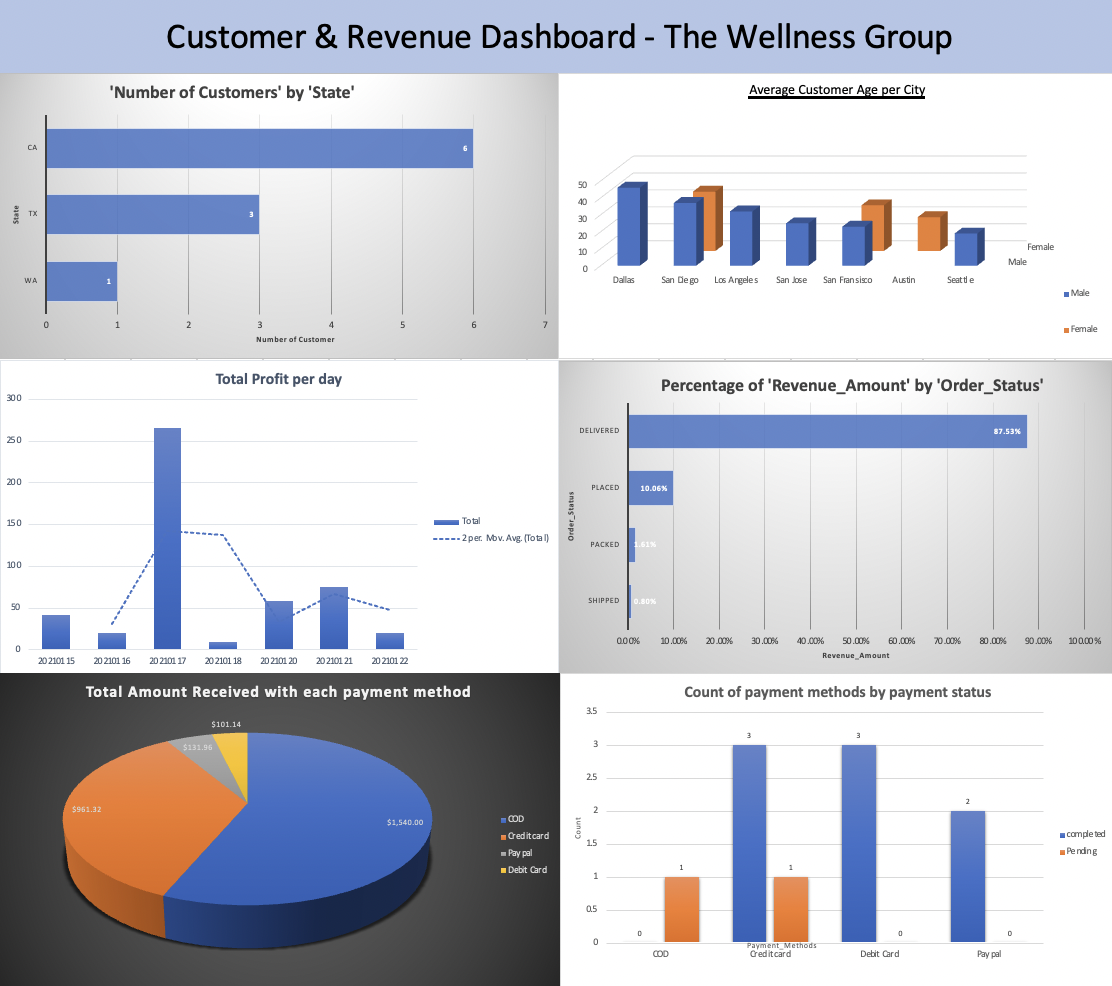
* 1. **Wellness\_Coach\_Incentive\_Fact:** This fact table is guided by the Primary Key **Incentive\_Key**. Contractor\_key, Product\_key, Purcahse\_Key and Time\_Key connect this fact table with their respective dimension tables to retrieve additional information about each measured fact. We will discuss these dimension tables individually. This fact table also includes Num\_of\_Prod\_Sold which tallies products sold for the agents. From this table, we can determine if the incentive was earned.
  2. **Contractor\_Dim:** This Dimension will link the incentive to the wellness coach who is contracted by Team Wellness. It contains their basic information such as name and department code but also their monthly sales goal which must be achieved to earn the bonus.
  3. **Critiques\_Dim:** Critiques are both complaints and recommendations that occur at specific points in time and sometimes belong to wellness coaches. They are tallied in the facts table and disqualify wellness coaches from earning the incentive if received on a quarterly basis, which is why they have an attribute Critique\_Date to capture their occurrence.
  4. **Purchase\_Dim:** Contains the Primary Key Purchase\_Key which is linked to a Foreign Key on the fact able Order\_Fact and Wellness\_Coach\_Incentve\_Fact. From this we can confirm completed purchases, their associated total amount and the date of purchase.
  5. **Order\_Fact:** The primary key of Order\_Fact table is Order\_Key. It also contains foreign keys such as Contractor\_Key, Purchase\_Key, Product\_Key, Critique\_Key and Time\_Key. Total\_Sales\_In\_Dollar and Total\_Profit\_in\_Dollar can be used as an additional validation source for both the criterions of the incentive, sales and complaints. It is also liked to the purchase dimension by Foreign Key.
  6. **Time\_Dim:** This dimension identifies events in time, particularly orders and critiques. This is important because the incentives are earned over a quarterly period therefore sales and complaints are tallied based on their occurrence during those time frames.

**DASHBOARDS:**

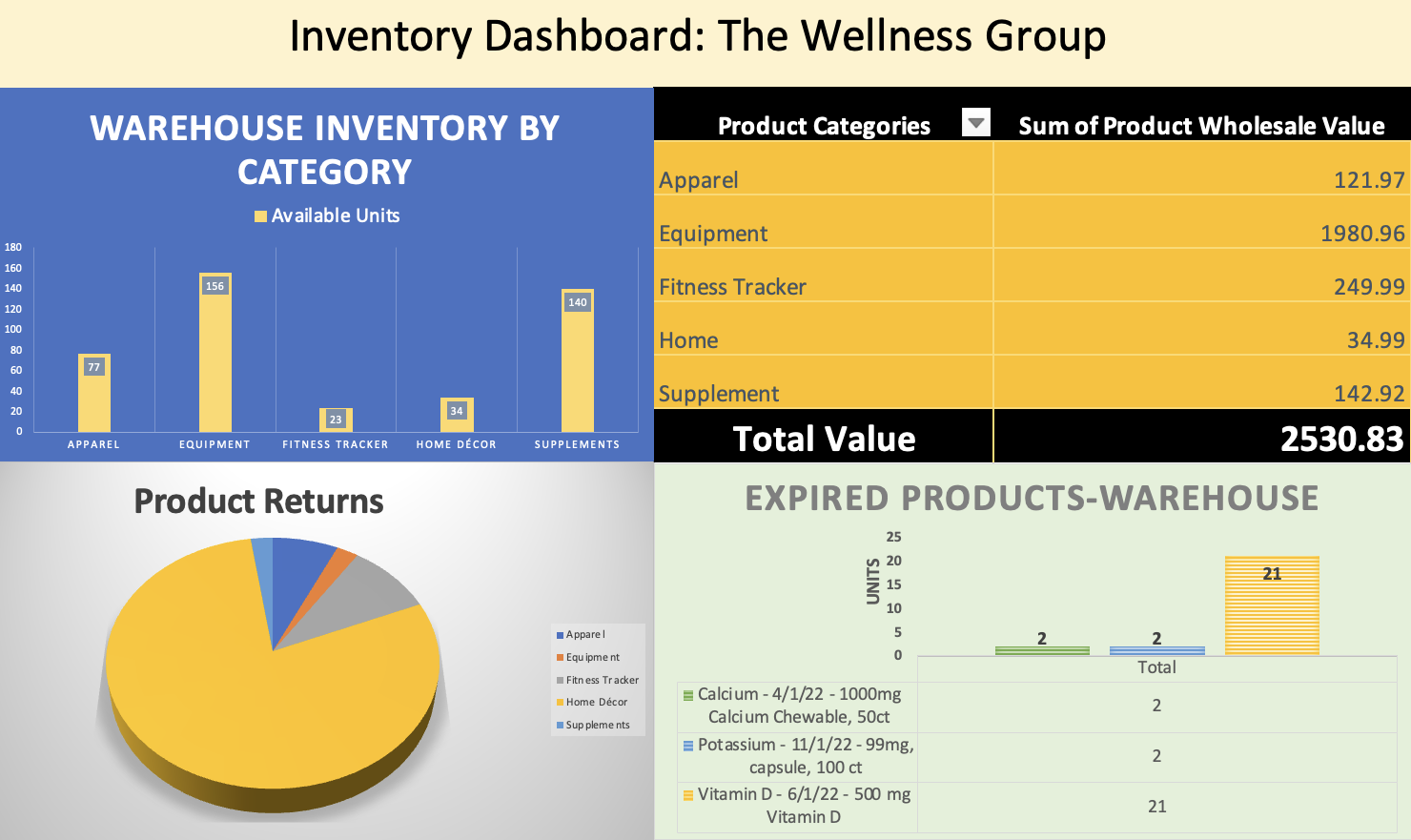
Our company will be using various dashboard modules to analyze company data and performance as efficiently as possible.

1. Our Customer and Revenue Dashboard will be used to monitor the locations, sex, and age of our customers. This customer information will assist our marketing team and help us decide where to locate future warehouse locations. This dashboard was created with a small amount of test data. We anticipate having approximately 10,000 unique customer accounts entered at our initial launch.

Revenue information is also on this view, so we can see our profit, percentage of revenue amount by order status, total amount received with each payment method and the count of payment methods by payment status.



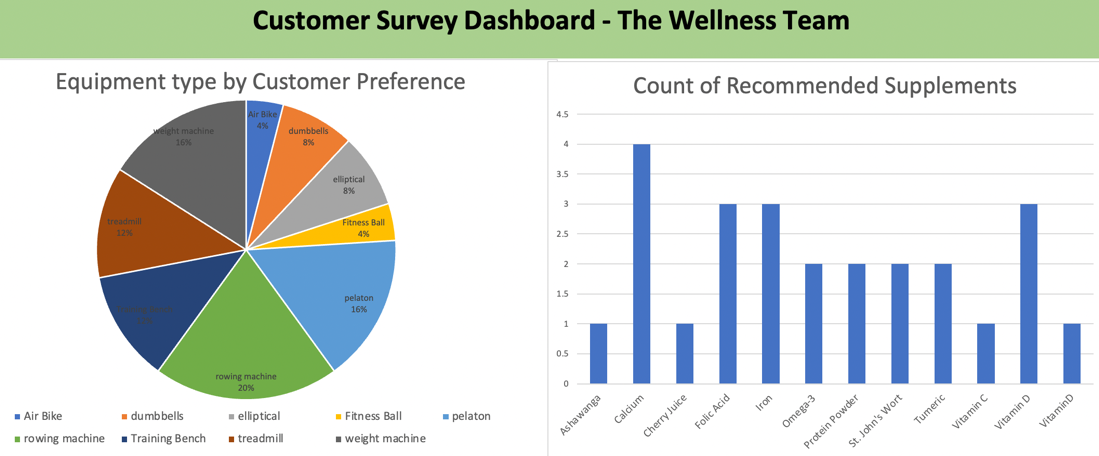
1. We will use an Inventory Dashboard to track our warehouse stock and monitor product expiration dates, as well as returns and total wholesale value of our warehouse stock. This way we can ensure we keep popular items in stock and monitor product quality.



1. Performance incentives are a big part of The Wellness Team’s concept, as we want to reward our wellness coaches for good performance. These dashboard tools will help us monitor wellness coach performance. The working concept is that the Wellness Coaches will be rewarded with a bonus for achieving their sales goal and not having any complaints against them. The dashboard view below displays the incentive payouts by quarter as well as complaints and recommendations breakdown. There is also a top performers section showing employees who have sold more than 300 products year to date.



1. Customer surveys are important to our business as they help us understand what services our customers want (for example, what kind of supplements or exercise equipment they are interested in) and offer valuable insights that can help us build our business.



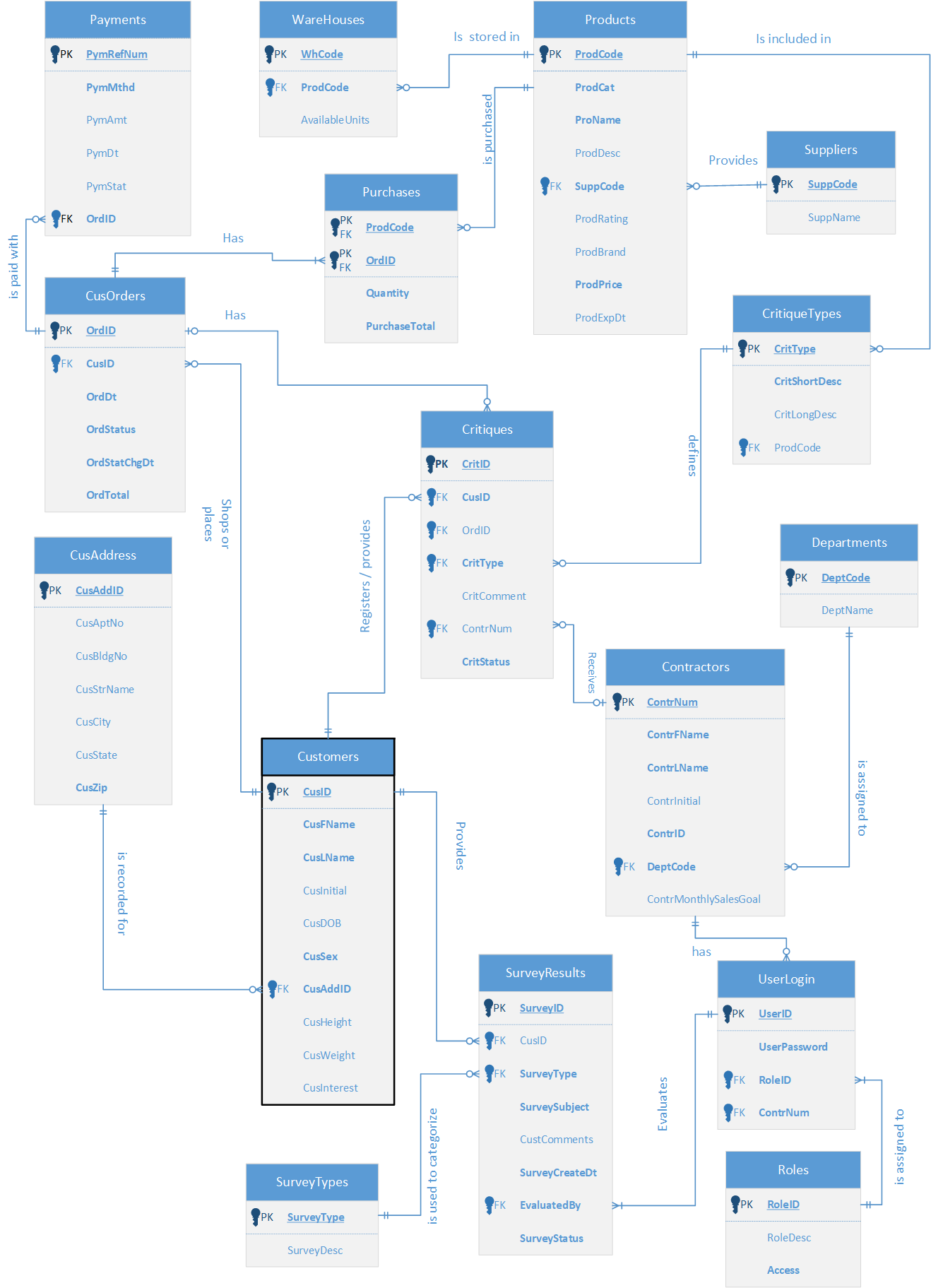
These are initial dashboard designs with a small amount of dummy data. Later we will adjust categories to allow for the most relevant and easy to reference categories. For instance, in the “Count of Recommended Supplements” chart, we will group similar supplements into groups such as “Vitamins”, “Juice”, “Protein Products”, “Performance Enhancers” and “Herbs”.

All dashboard designs are subject to revision. Thank you for your consideration.

**References:**

1. Michał Drozdowicz (2008). Designing And Implementing Data Mart For An Agent-Based E-Commerce System. *IADIS International Journal.* Retrieved from <https://nationalu.brightspace.com/content/enforced/32006-ANA655-35752-2206/IADIS_2008_data_mart.pdf>

APPENDIX 1: ERD DIAGRAM



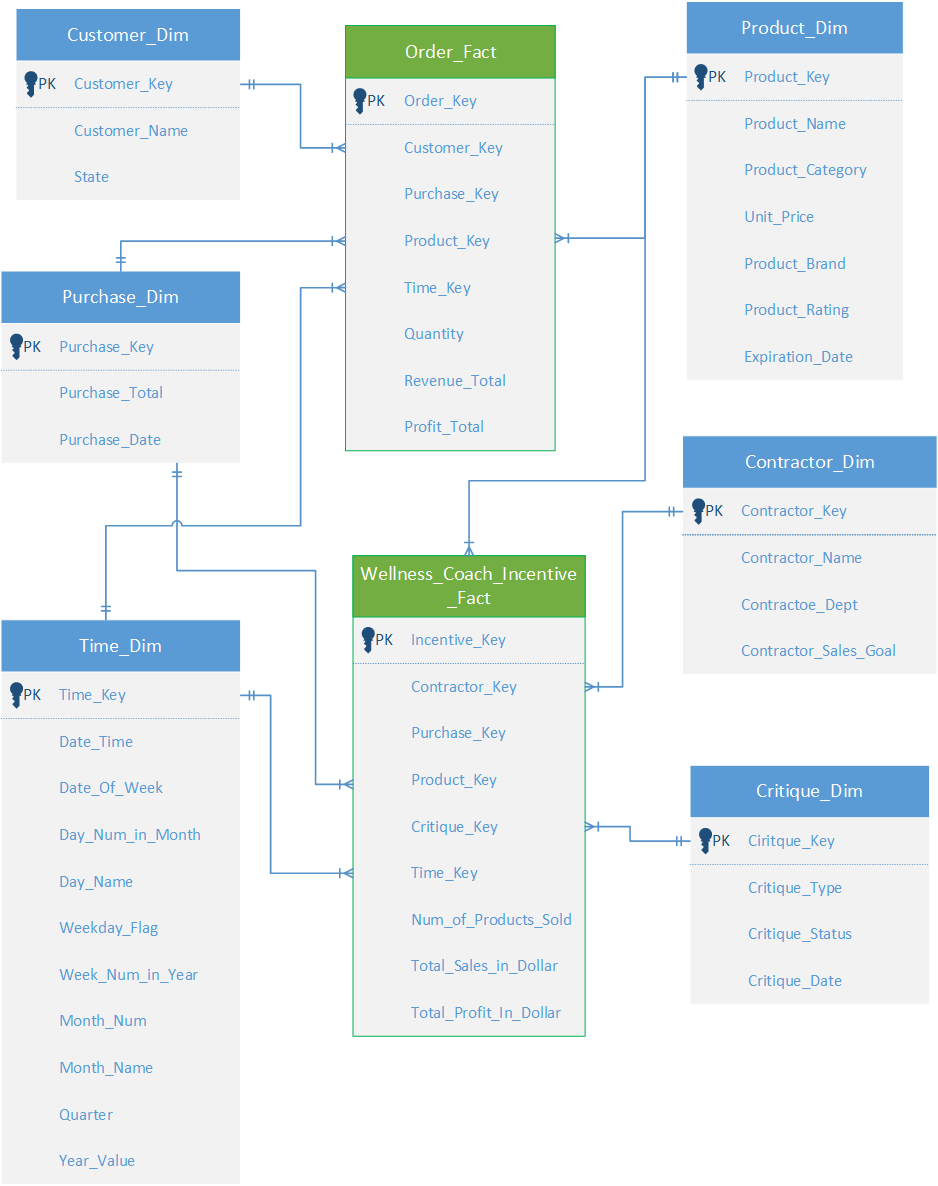
APPENDIX 2: DATAMART STAR SCHEMA

Graphical user interface

Description automatically generated

APPENDIX 3: WELLNESS COACH INCENTIVE SCHEMA

Modified this star schema as per evaluator’s suggestions. Purchase\_Dim is kept dimension which contains the total amount and date associated with each purchase. It is connected to order\_fact and wellness\_coach\_incentive\_fact tables with foreign key **Purchase\_Key.** And measures of orders are stored in Order\_Fact table, which is also connected to Customer\_Dim.



1. Payment transactions imported into staging from an Ops source are a good candidates to be migrated into a data mart a a fact table because metrics such as payment\_amt can be totaled up for a period of payment completed. A fact table for purchases would normally be joined to Customer (at least for account level summary kinds of queries, or to aggregated for demographics of similar customers)… and perhaps for visibility of rejected payments for any reasons that are of interest, such as account closed or insufficient funds. The the loss of merchandised would be written off as a normal practice if it is not returned. If A Purchase fact is set up, the attributed Payment Method could be denormalized into instead of having a new dimension table, . [↑](#footnote-ref-2)
2. Normally in the business world, the people in operational support roles would handle payments that failed to become complete as a funds transfer from a bank. This may happen after payments initiated in the ops source are forwarded onward to the financial systems where people who know what to do take action, to collect the money or to lock out a customer so that the fraud would not be repeated. [↑](#footnote-ref-3)