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,

June29 earned 45.

Many edits in green provided to suggest improvements that can result in this part3

being improved & resubmitted for full score + a bonus.

Project - Week 4 - Part 3

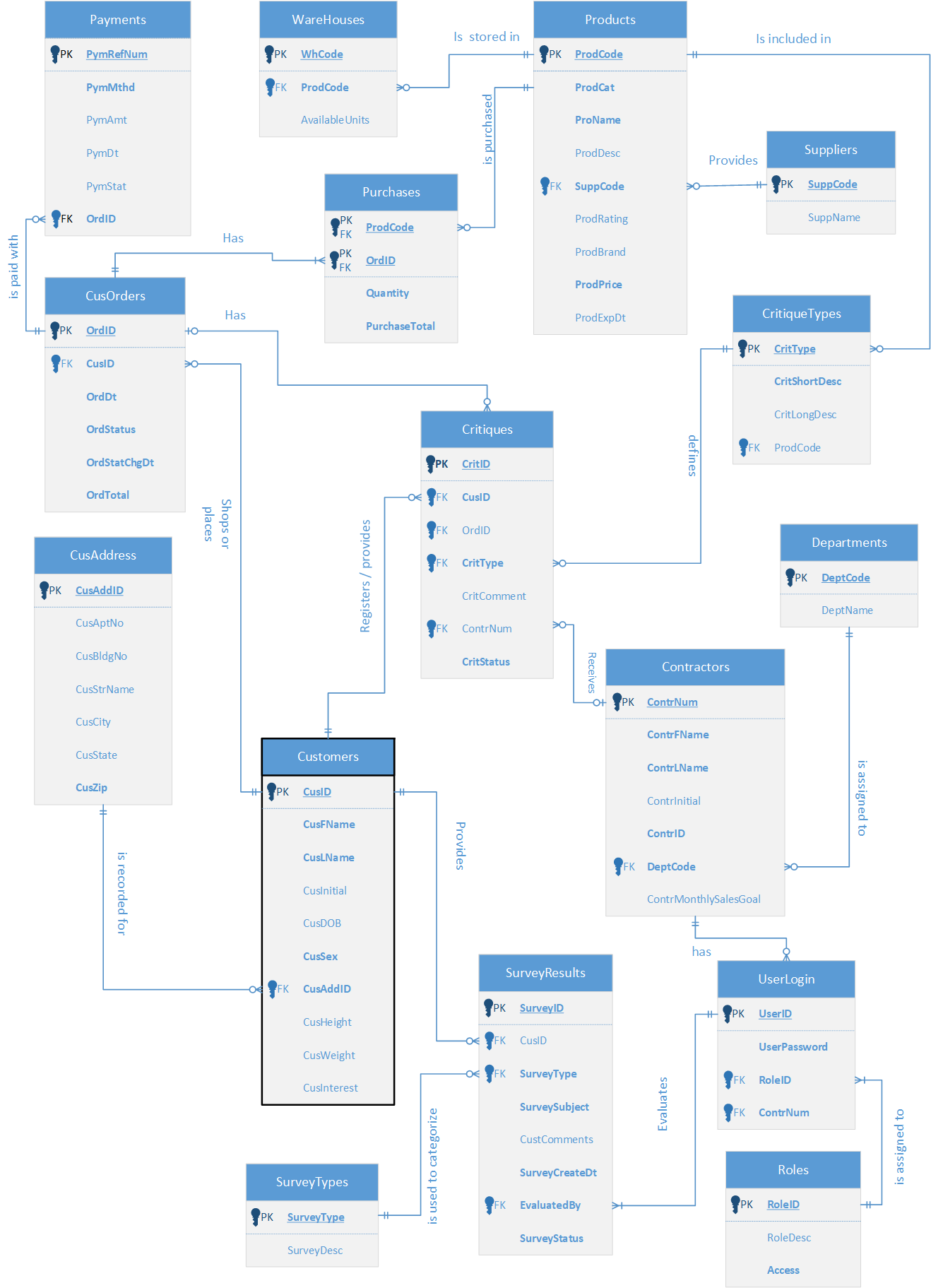
June 24th, 2022

**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:

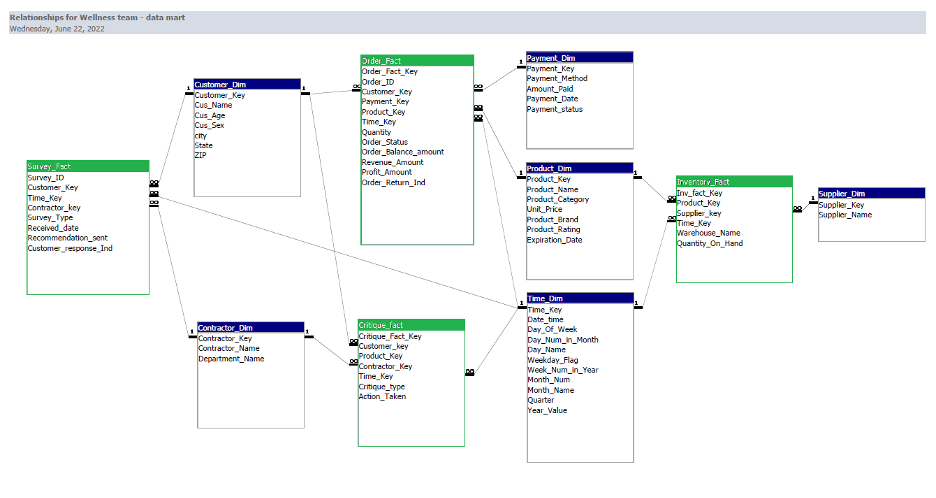
1. Modified name of all entities to be plural e.g., Products, Employees, CusOrders.
2. Added new attributes in Purchases table – Quantity and PurchaseTotal.
3. Added new attributes to Products table – ProdBrand and ProdPrice
4. Added new attribute to CusOrders table – OrderTotal.
5. Removed ‘PymMthd’ 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 that we propose to build <design> [[[1]](#footnote-2)] in this project. There will be a separate pdf file attached showing the same star schema along with an access file that will contain a table design view and schema design.



The scheme was created by The Wellness Team. We referred to a paper by M. **Drozdowicz** provided by instructor Dr. Barbara Lauridsen. *A larger version is included in the appendix.*

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. *Evaluator’s note: this entire structure may best be an extension of the original Ops database. I am unconvinced that a data mart is the home for survey results.*
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 become> transformed and loaded into this table to <provide> measure metrics such as “Total Number of orders in each status (placed [[[2]](#footnote-3)]/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 open [[[3]](#footnote-4)] critiques (such as complaints), 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 with to primary keys of Customer, Product, Contractor, and Time dimensions <for purpose of searching history for patterns of activity seeking possible resolutions>.
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 > of a customers such as Customer Name, Age, Sex from Customers table and City, State, Zip //from CusAddress/[[4]](#footnote-5)/ table in operational source. 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. and 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> the number of contractors <assigned> in each department.
3. **Payment\_Dim:** 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 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.

Diagram, schematic

Description automatically generated

*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:

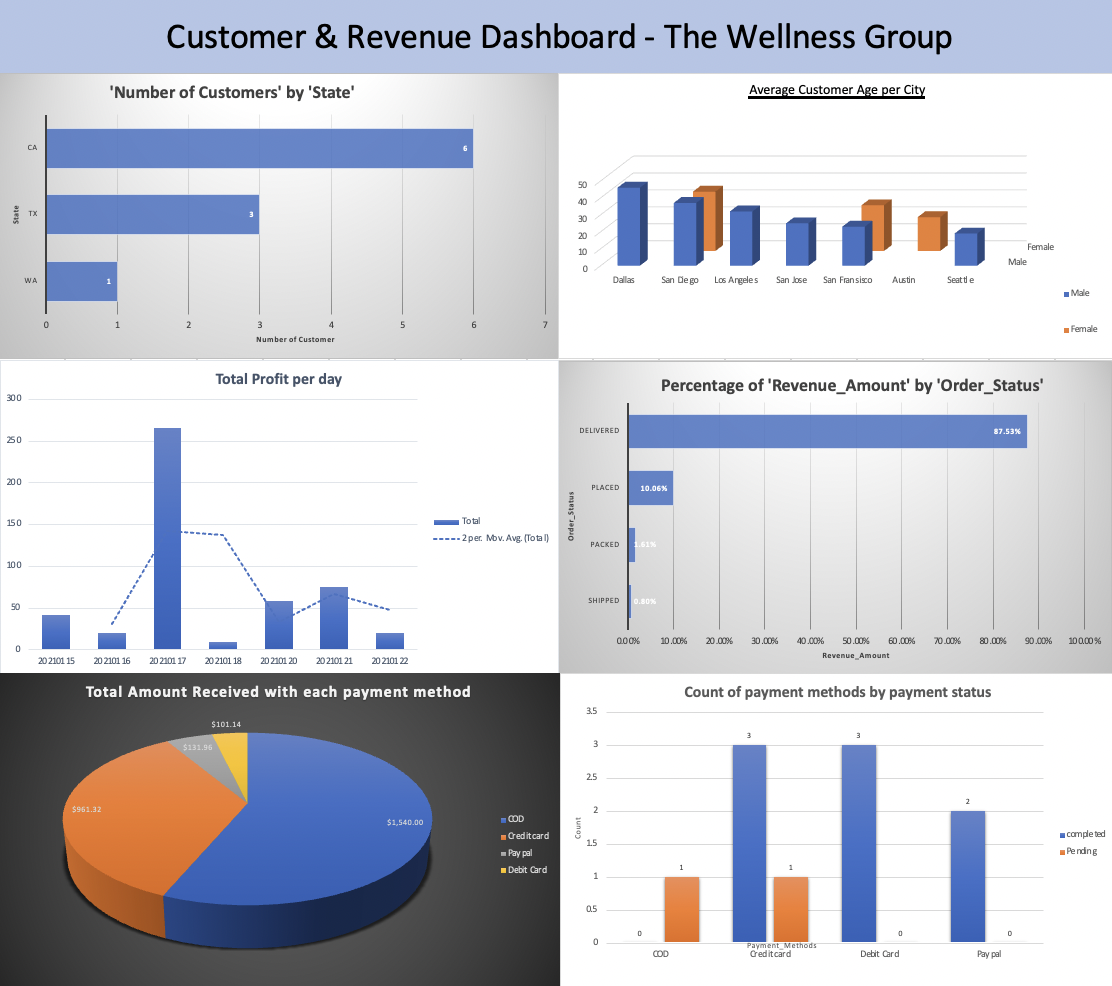
* 1. **WellnessCoachIncentivesFacts:** This fact table is guided by the Primary Key ContrNum or Contractor Number which is linked to our Wellness Coach’s. Unique attributes in this dimension include CritTotal and RecommendTotal which are tallies of these critique types. There is also ProdTotalSold which tallies products sold for the agents. From there, the Foreign Keys, CritiID, OrderID, ProdCode and Time\_Key link to our dimension tables which will be discussed individually. From this table, we can determine if the incentive was earned.
  2. **ContractorDim:** 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. **CritiquesDim:** 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 are linked to the time dimension, to capture their occurrence. They are linked by Foreign Key to the Contractors dimension as well.
  4. **PurchaseDim:** Contains the Primary Key ProdCode which is linked to a Foreign Key on the fact able. From this we can confirm completed orders and their associated products.
  5. **CusOrderDim:** Is the order dimension which contains information such as order data, date, and status. It is linked by Foreign Key to the Time and Critique's dimensions so it 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. **TimeDim:** 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:**

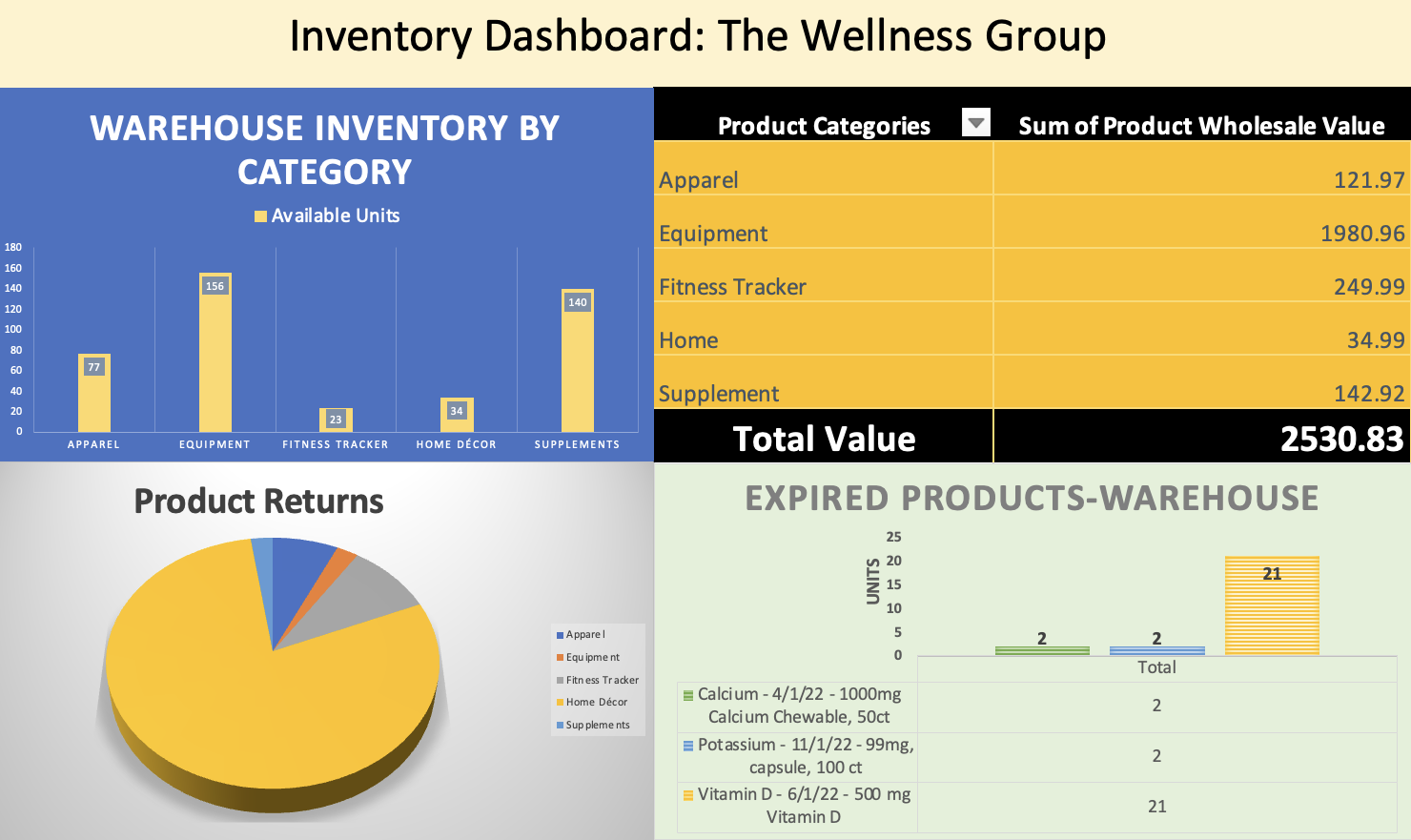
Our company will be using various dashboard modules to analyze company data and performance as efficiently as possible.

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.



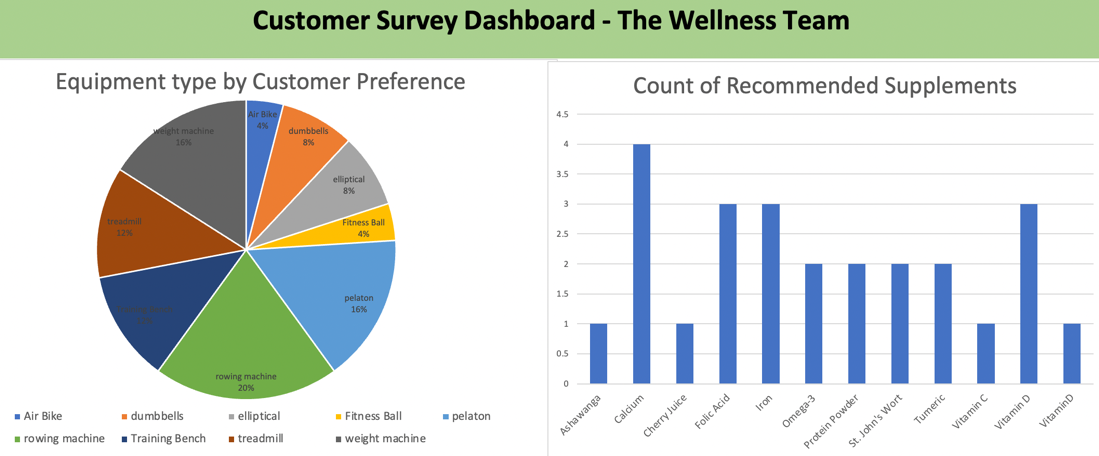
1. We will use an Inventory Dashboard to track our warehouse stock and monitor product expiration dates, as well as returns and total wholesale value of our warehouse stock. This way we can ensure we keep popular items in stock and monitor product quality.



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.



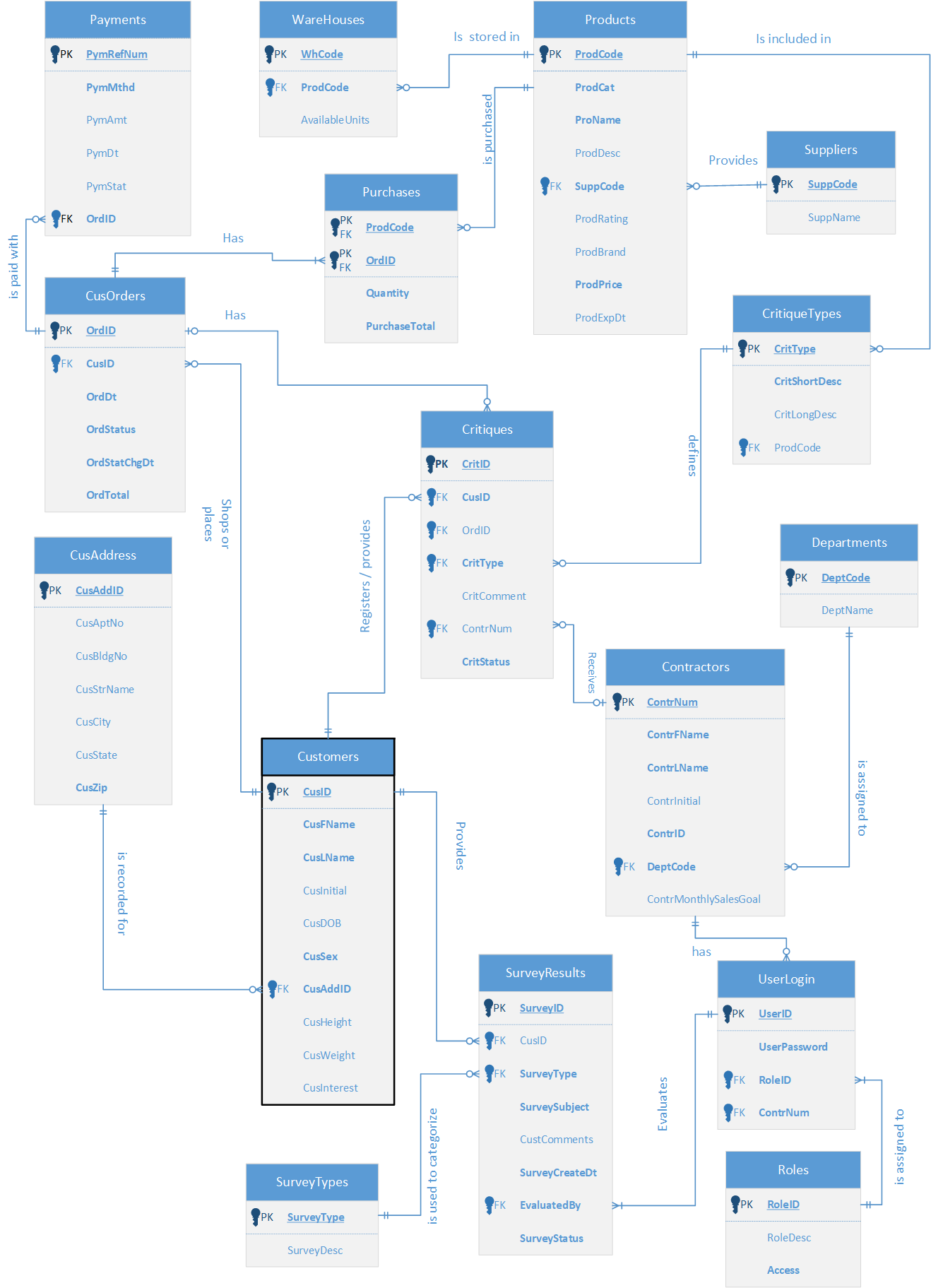
These are initial dashboard designs with a small amount of dummy data. Later we will adjust categories to allow for the most relevant and easy to reference categories. For instance, in the “Count of Recommended Supplements” chart, we will group similar supplements into groups such as “Vitamins”, “Juice”,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

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

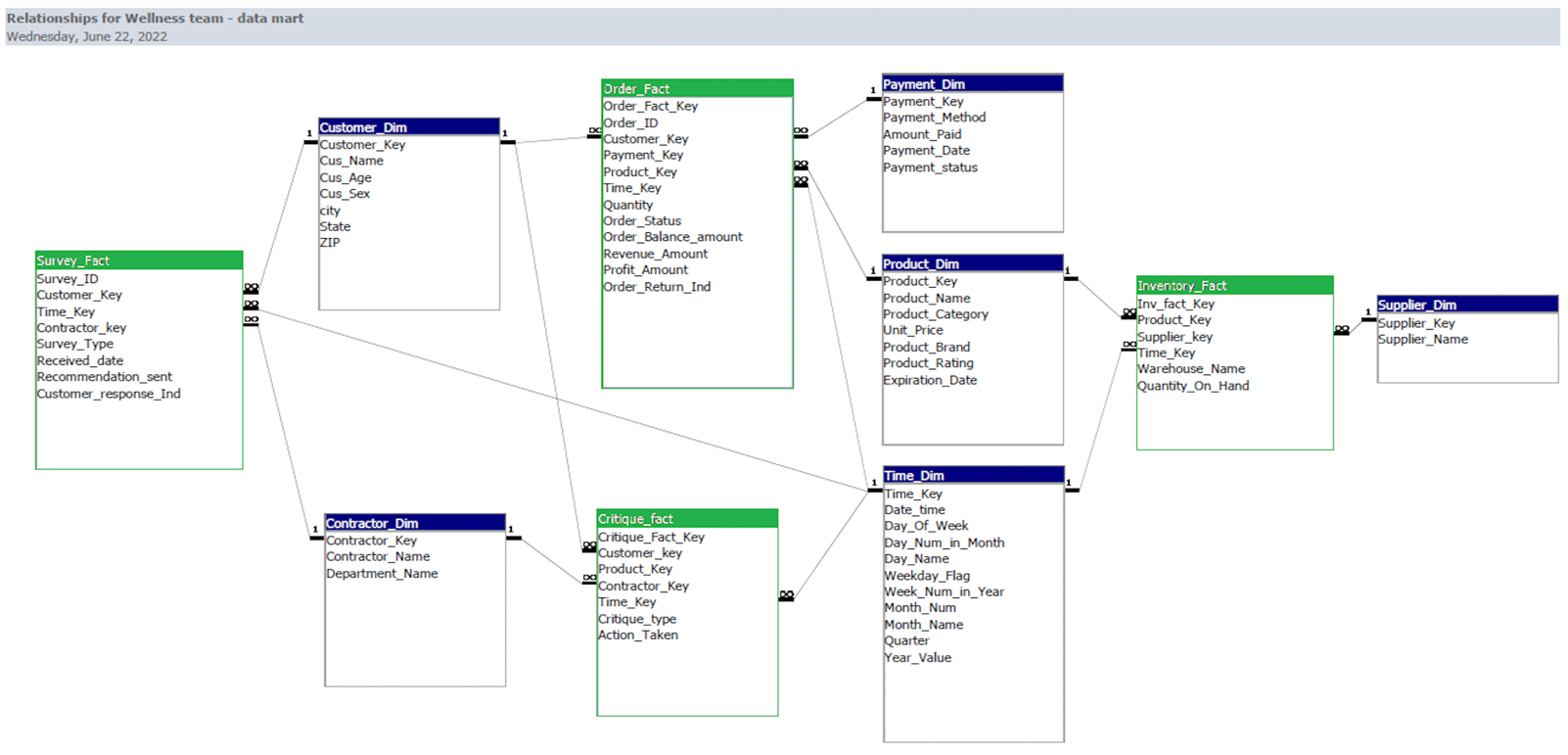
**References:**

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



APPENDIX 3: WELLNESS COACH INCENTIVE SCHEMA

Evaluation note: Relationships drawn from a dimension primary key are always for a 1 side of a 1-to-M, wereas the connection to a fact table’s foreign keys will have the crow’s foot for the many side. If a dimension needs a date, the attribute is just a date… not connected to the time dimension. Upper left entity PurchaseDim is wrong since it contains measures; can be renamed to be Purchase\_History. It will need to be linked to Customer\_Dim. This model is a reasonable sketch for the ways that SQL queries may be designed to track productivity of a coach.

Diagram, schematic

Description automatically generated

1. ANA655 scope is not to actually build the designed pilot dimension model. [↑](#footnote-ref-2)
2. New orders placed into the source Operations database will linger there until they are ready for migration. When no longer volatile (being in progress for fulfillment) then it makes sense to extract them into the data mart. Tracking delivered shipments is often a reason to bring historical transactions to architecture’s staging. A status of “returned” is a reason to import transactions that start out in the Ops source so that revenue for the original order can be reversed out, and perhaps evaluated for a reason why a customer did not keep it. [↑](#footnote-ref-3)
3. We can acknowledge that handling of “open” survey results will often remain a responsibility within the Ops system… handled by people who know the situation. A reason to import them into a data mart would be whenever a lot of them need to be evaluated to make decisions to correct a problem, or to train someone to handle them. However, if a long history of critiques exists, such as on a spreadsheet, then perhaps importing them into a data mart in order to use OLAP tools to view the patterns that can be mapped to product known in the dimension model, then that is acceptable. [↑](#footnote-ref-4)
4. Some attributes in the Ops sources should omitted from the design of a dimension based on how facts will be queries for making business decisions. [↑](#footnote-ref-5)