Về phân tích theo từng transactions thì ta thấy được số lượng của mỗi transactions sau khi mua tập trung chủ yếu quanh mức 10-12 item/ 1 transactions là rất cao => shop bán được rất nhiều hàng khi khách ghé tới mua.

Sau đó thì em group từng mặt hàng lại và đếm số lượng giao dịch trên mỗi mặt hàng và plot nó lên thể hiện theo %. Thông qua biểu đồ này, ta thấy được số lượng mua của khách đối với các mặt hàng đều rất cao, trên mức 30% và cao nhất là chuối và phô mai, thịt xông khói.

Em phân loại các loại items theo 5 categories khác nhau và thứ tự của từng item ứng với % bán ra được cao nhất, ví dụ về fruit ta có banana bán ra được cao nhất nên banana sẽ ở vị trí 1, tương tự Fatty ta có cheese và food ta có bacon. Nhờ vào sự sắp xếp này ta sẽ biết được nên để ưu tiên sản phẩm nào để khi khách hàng bước vào họ có xu hướng mua ngay lập tức.

Tiếp đến từ mlxtend.patterns, em import apriori và association rules vào trong môi trường colab. Đầu tiên, sẽ cho chạy apriori và để min support=0.2 để mục đích cho thuật toán tính rằng mức hỗ trợ đạt 20% thì mới công nhận, nghĩa là mặt hàng item nào hoặc cặp itemsets nào được khách hàng thực hiện trong transactions từ data trên 20% thì sẽ trả ra kết quả đúng. Sau đó, mô hình sẽ trả ra kết quả có dãy data là frequently\_bought\_items có 66 itemsets được khách hàng mua thường xuyên trên 20%

Tiếp theo chạy association rules, lấy data input vào là frequently\_bought\_item vừa mới chạy ra kết quả ở phía trên, và ta sẽ để ràng buộc cho mô hình là confidence với mức hỗ trợ đạt phải trên hoặc bằng 50%. Sau đó ta có được kết quả là 45 cặp item sets : antecedent: mặt hàng khách hàng mua đầu tiên. consequent: mặt hàng khách hàng mua tiếp sau khi họ mua antecedents antecedent support: đo lường số lần khách hàng mua mặt hàng antecedents xuất hiện trong toàn bộ dữ liệu giao dịch consequnet support: đo lường số lần khách hàng mua mặt hàng consequents xuất hiện trong toàn bộ dữ liệu giao dịch support: mức độ mua hàng thường của khách hàng đối với mỗi cặp antecedents và consequents. ví dụ hàng 0 có support= 0.211207-> 21,12% cặp bread và banana sẽ được khách hàng mua khi ghé cửa hàng confidence: mức độ tin cậy đo lường. Ví dụ hàng thứ 0 có cặp bread và banana thì cofidence= 0.518519 -> khách hàng mua bread thì 51,85% họ sẽ mua banana

Để trực quan hóa dữ liệu 1 cách rõ ràng hơn, em sẽ sử dụng đến thư viện networks để tạo đồ thị liên kết lấy mỗi nhánh là các tập antecedent và consequent và confidence thể hiện mức độ tương quan chặt chẽ giữa antecedent và consequent. Kết quả sẽ trả cho ta ra được hình như bên phía dưới.

Sau khi chạy ra được kết quả, e vẽ lại để nhìn rõ hơn về sự liên kết giữa các itemset. Ở đây, ta thấy phần lớn mặt hàng sẽ tập trung ở banana, cheese, bacon, và 3 mặt hàng này cũng đứng đầu trong từng categories có % tỉ lệ bán hàng cao nhất, có nghĩa là người mua hàng khi người ta mua 1 sản phẩm trước đó bất cứ loại gì thì phần lớn mặt hàng thứ 2 họ mua sẽ nhắm tới 3 item tiếp theo là banana, cheese hoặc bacon. Salt và shaving foam cũng đứng đầu trong % tỉ lệ bán ra cao của mỗi categories nhưng mức confidence liên kết cũng không được nhiều, ở đây thể hiện rằng chỉ khi khách hàng mua bread thì 50% họ sẽ mua salt, và ở đây thì thể hiện khách hàng mua bacon thì 50% họ sẽ mua shaving foam.

Thông qua mô hình liên kết phía trên, nhóm dựa vào đó để sắp xếp thứ tự các mặt hàng để phù hơp với nhu cầu mua hàng đến từ khách hàng để tối ưu hóa không gian, và thời gian giúp cho khách hàng thuận tiện hơn trong việc mua sắm của họ. Dựa trên 22 mặt hàng, nhóm chia thành 6 categories chính và mỗi categories sẽ bao gồm các itemset liên quan mật thiết với nhau. Các loại mặt hàng về trái cây và các sản phẩm về sữa có thời hạn sử dụng gấp và tránh tồn kho nên nhóm sẽ đặt fruit và milk product ở trước cổng vào, banana và cheese là 2 mặt hàng có tỉ lệ bán ra được cao nhất nên 2 sản phẩm này sẽ được xếp lên kệ đầu tiên. Dựa thêm vào confidence, nhóm cũng sẽ xếp ưu tiên những items hoặc kệ nào đặt liền kề với nhau, ví dụ ở food chúng ta có bread thì do là khi người ta mua bread thì 50% sẽ mua salt và ngược lại, mua bread thì sẽ mua cheese cho nên qua hình 3 kệ này được đặt sát nhau. Ngoài ra, sản phẩm butter, khi khách hàng mua butter thì 50% sẽ mua cheese hoặc bacon cho nên 3 kệ cũng được đặt cạnh gần nhau. Tương tự, mua Cheese thì sẽ mua egg và mua egg sẽ mua honey. Do là khi khách hàng mua bột flour thì họ chỉ mua duy nhất hazelnut là hạnh nhân thì nhóm nghĩ là khách hàng sẽ có xu hướng mua 2 items này về để làm bánh cho nên lập 1 quầy bakery để 2 sản phẩm này gần nhau, và khách hàng mua heavy cream bột keo béo thuộc nhóm sản phẩm sữa thì 50 % sẽ mua hazelnut cho nên xếp quầy bakery gần lại với milk product sẽ rất thuận tiện cho khách hàng nào có nhu cầu tìm kiếm những items để về làm bánh.