

**SeUng Jung**

**Data Life Cycle on AWS**

*Lab 4. Visualization with Quicksight*

**Table of Contents**

Lab 설명 3

Lab Architecture 3

Visualization with Quicksight 4

# Lab 설명

이번 Lab은 Data Lake에 저장된 데이터의 분석 결과를 시각화 하는 단계입니다.

시각화의 목적에 따라 다양한 솔루션을 선택 가능합니다. Business Intelligence를 위해서는 AWS의 Quicksight 솔루션을 사용하여 다양한 접근을 통해 인사이트를 찾아냅니다.

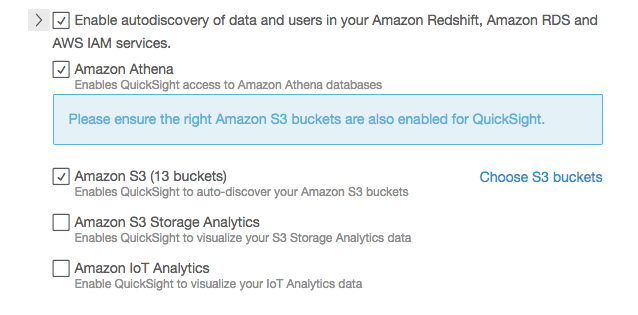
Dashboard를 구성하여 지표를 시각화 하고 공유, 분석 하기 위해서는 Zeppelin 웹 노트북을 사용합니다.

# Lab Architecture

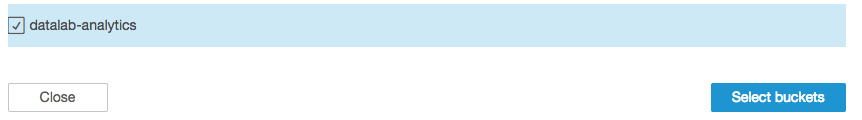


# Visualization with Quicksight

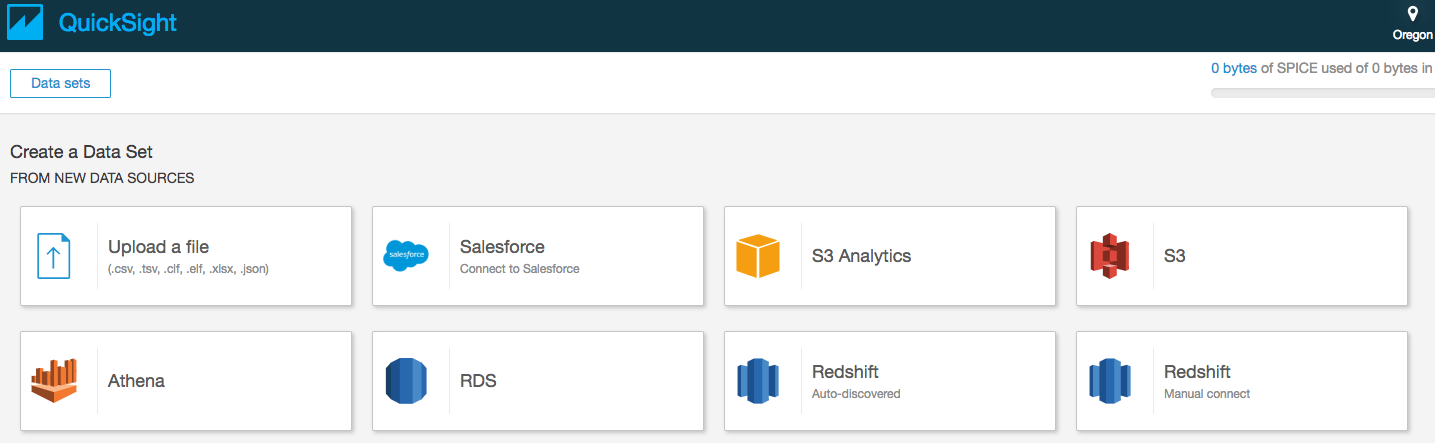
1. 우측 상단의 리전을 **Oregon (US-WEST)**으로 선택합니다.
2. AWS Management Console에서 Quicksight서비스 화면으로 이동합니다.
3. 최초 접속시에 **[Sign up for Quicksight], Standard edition**을 선택하여 account를 생성합니다.
4. 필요한 정보를 입력하고, Region 은 US West(Oregon)을 선택하고 다음과 같이 체크박스를 선택합니다.



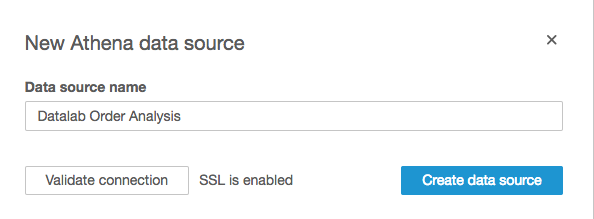
1. **[Choose S3 buckets]**를 선택하여 분석용 데이터가 들어있는 버킷을 선택합니다.



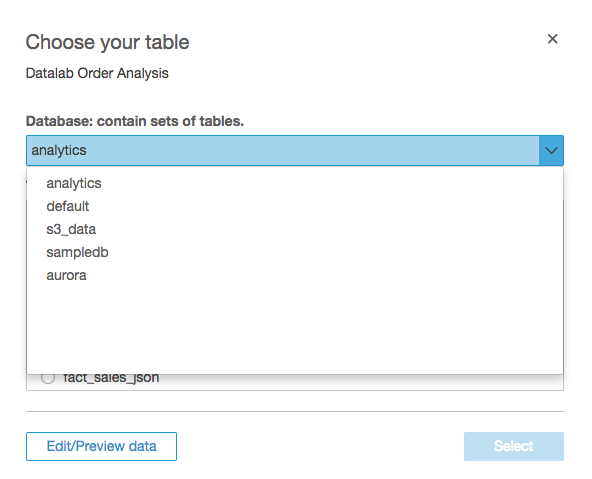
1. 좌상단의 **[New analysis], [New Data set]**을 선택하고, **Athena**를 선택합니다.



1. Data source name에 **Datalab Order Analysis**을 입력하고 **[Create data source]**를 선택

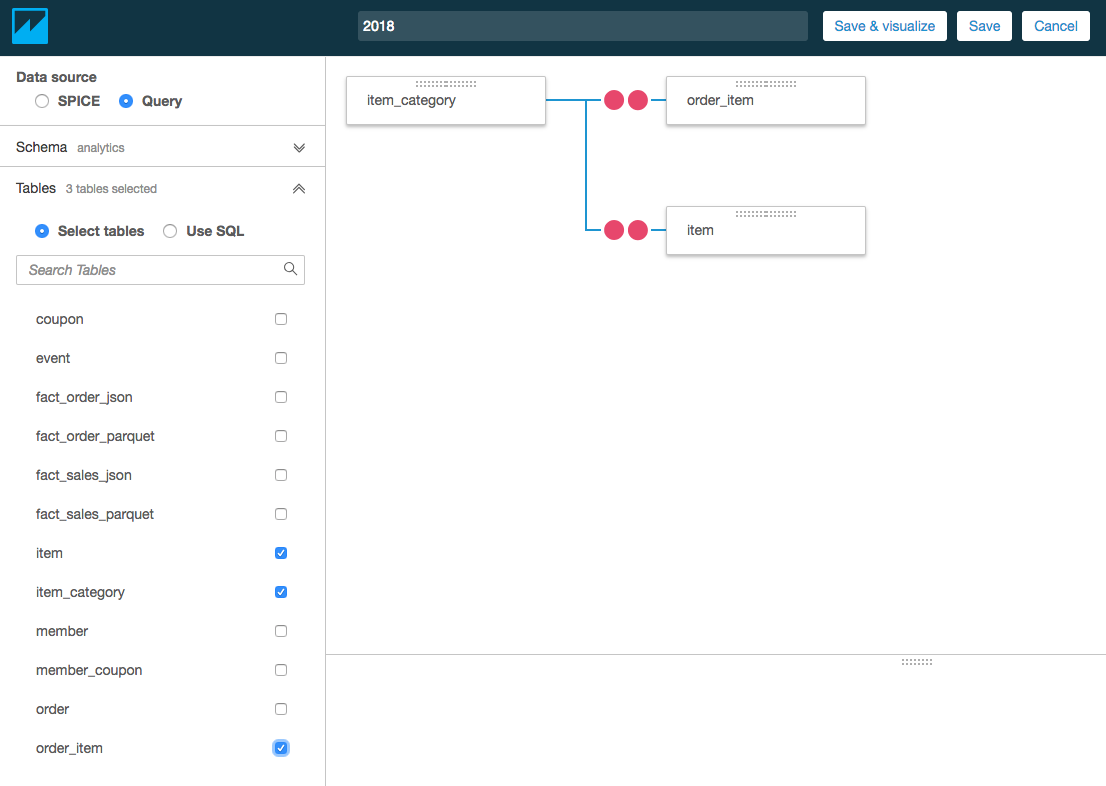


1. Database에서 기존에 만들어진 Data Catalog와 동일한 목록을 확인할 수 있습니다. Analytics를 선택하고 좌측의 Edit/Preview data를 선택합니다.

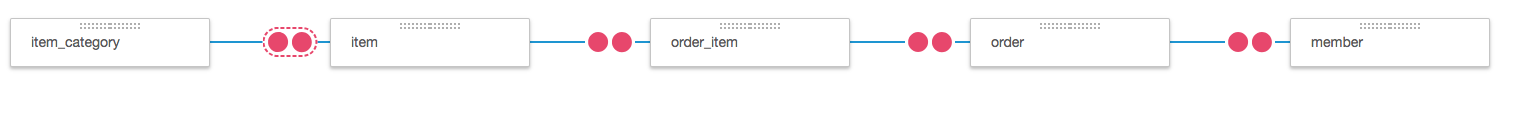


1. 다음처럼 분석에 사용할 테이블을 선택합니다.

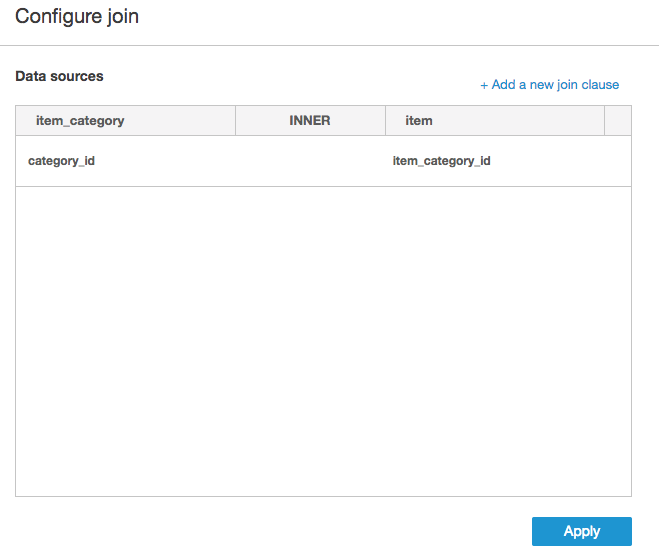
Schema : **analytics**, Tables : **item\_cetegory**, **item, order\_item**



1. 각 테이블 상자는 상단의 ……. 부분을 클릭하여 조인 대상 테이블 위로 옮기면 연결선이 변경됩니다. 다음처럼 연결합니다.



1. 첫번때 **item\_category**와 **item** 사이의 연결을 클릭하고 아래와 같이 키값 맵핑을 선택합니다. 그리고 **[Apply]**를 눌러 저장합니다.



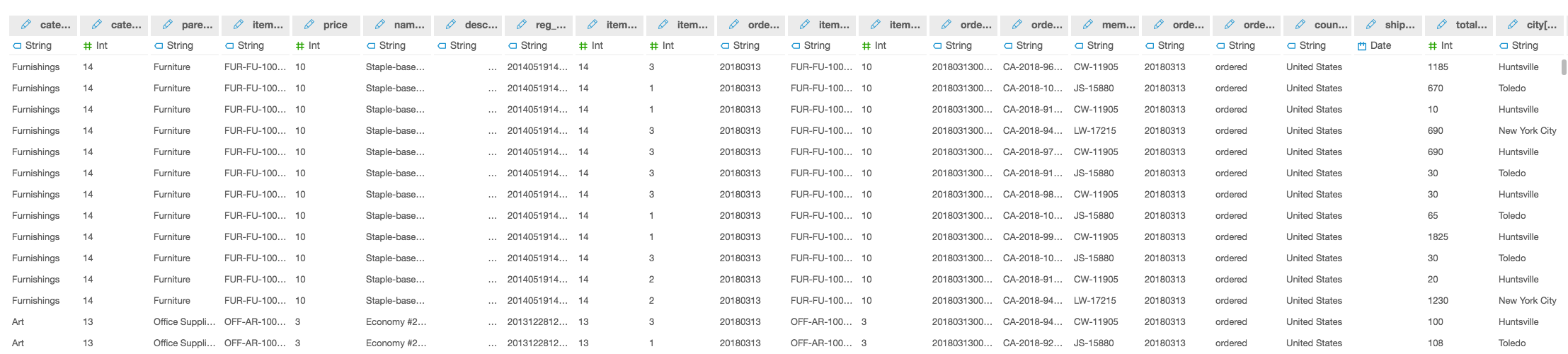
1. 순서대로 다음과 같이 설정합니다. 각각 **[Apply]**를 눌러 저장합니다.



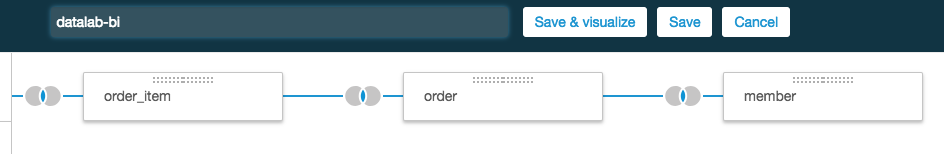




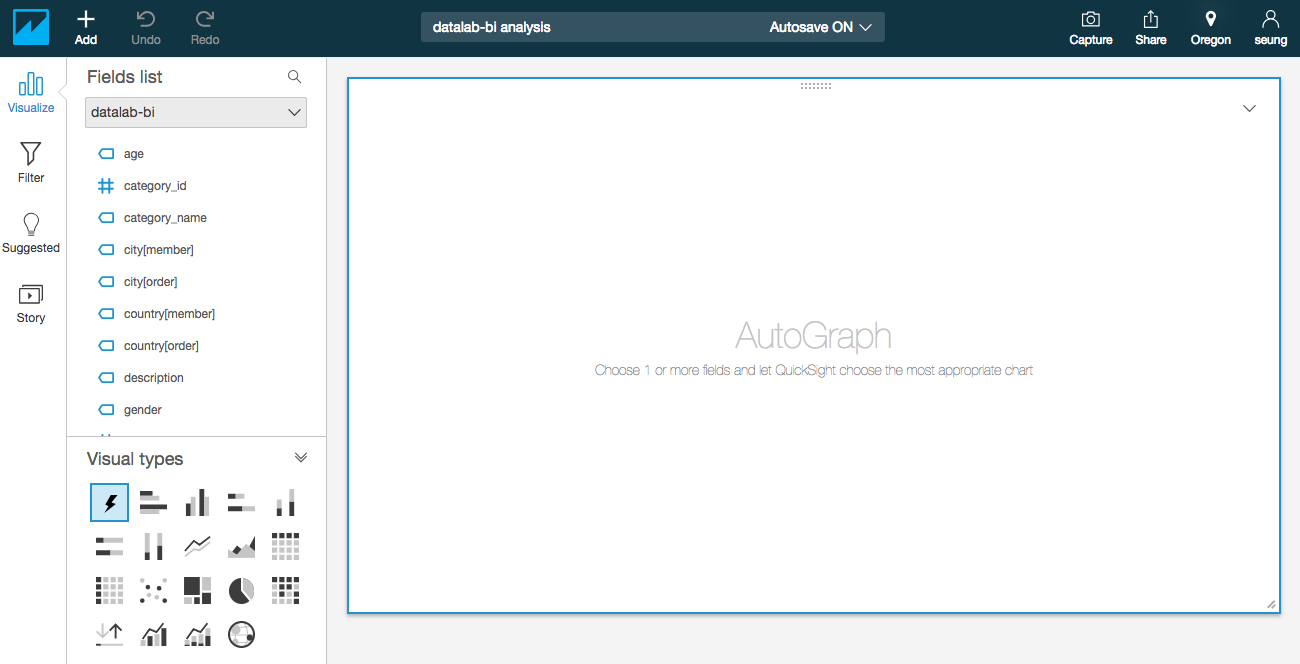
1. 테이블간 키 맵핑이 완료되면 다음처럼 데이터가 확인 됩니다.



1. 상단에 데이터셋 이름으로 **datalab-bi**을 입력하고 상단의 **[Save & Visualize]** 버튼을 클릭합니다.



1. BI를 위한 모든 준비가 완료되었습니다. 좌측 필드 리스트에서 화살표는 코드성 정보인 Dimension을 나타내고, #는 값을 의미하는 Measure를 나타냅니다.



1. 첫번째 그래프를 그려봅니다. 카테고리별 성별에 따른 매출 규모를 비교해 봅니다.

Visual types : **Horizontal bar chart**

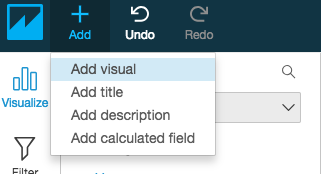
Dimension : **category\_name**

Measure : **total\_price**

Group/Color : **gender**



1. 두번째 그래프를 위해서는 좌상단의 **[Add visual]**을 선택합니다.



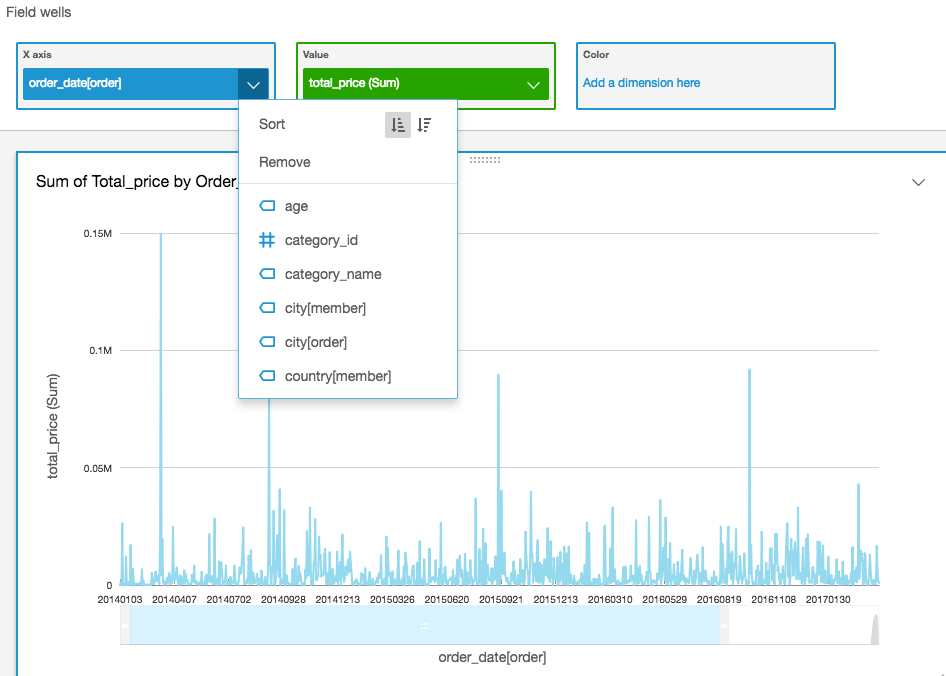
1. 두번째 그래프를 그려봅니다. 날짜별 매출 추이입니다.

Visual types : **Line chart**

Dimension : **order\_date[order]**

Measure : **total\_price**

X axis의 정렬을 내림차순으로 하고, 아래 범위를 넓게 설정하면 다음과 같은 뷰를 볼 수 있습니다.



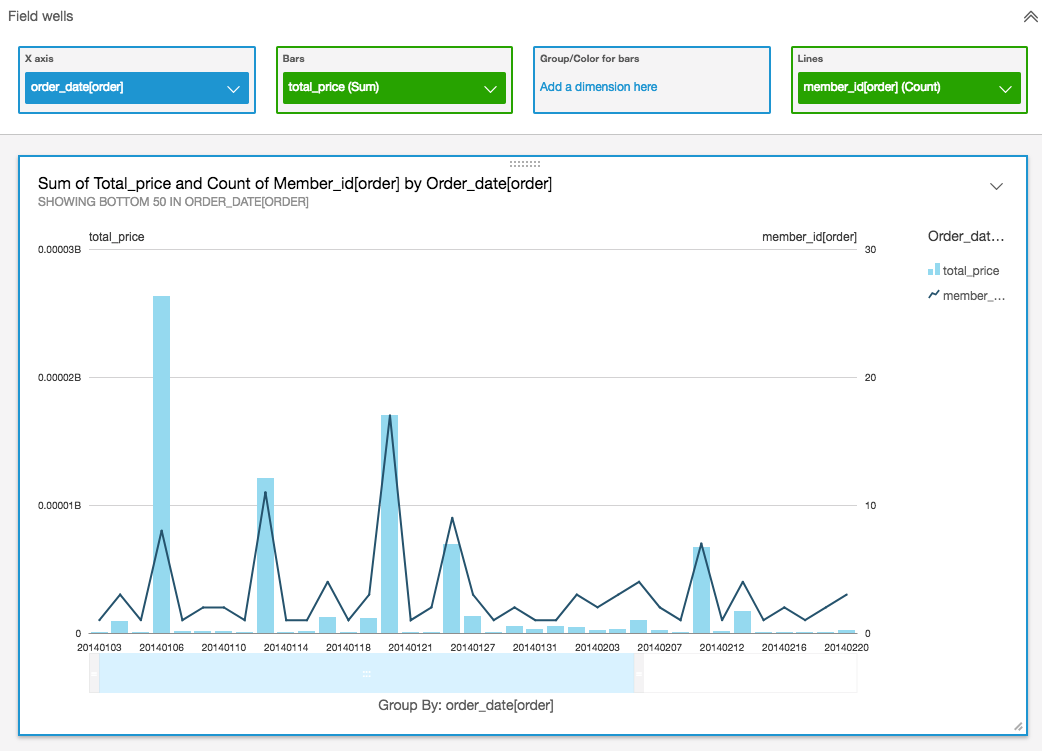
1. 세번째 그래프를 그려봅니다. 날짜별 매출과 구매 회원수 간의 상관관계를 확인할 수 있는 그래프입니다.

Visual types : **Clustered bar combo chart**

Dimension : **order\_date[order]** –> 정렬은 내림차순

Measure : **total\_price**

Lines : **member\_id** -> count 함수 선택



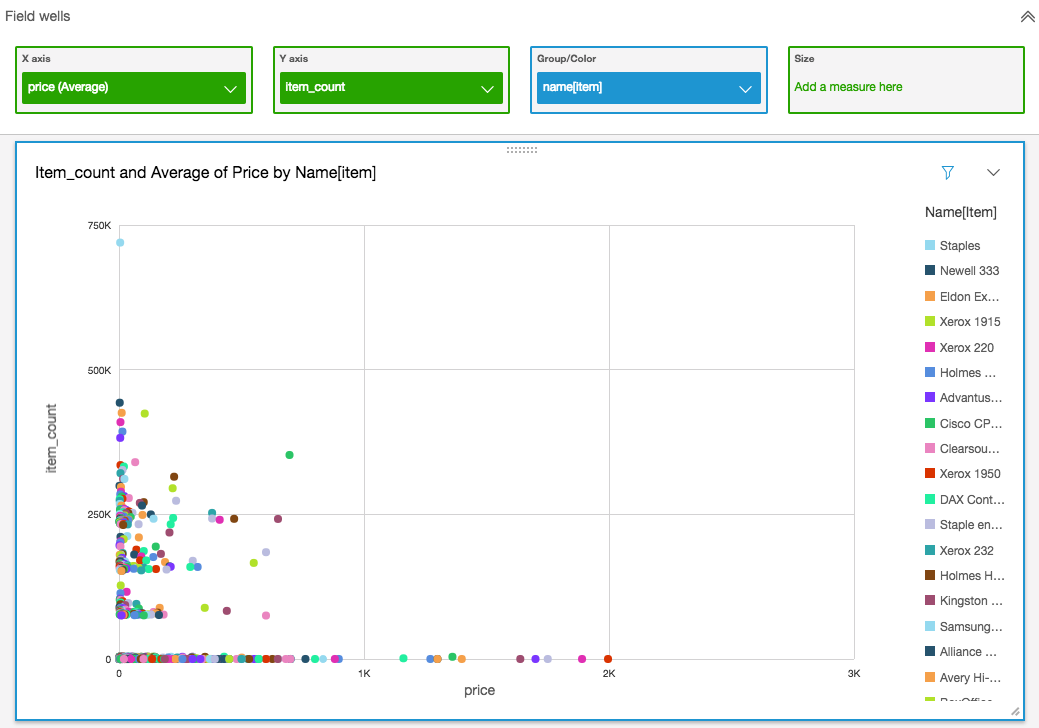
1. 네번째 그래프를 추가합니다.

Visual types : **Scatter**

X axis : **price(Average)**

Y axis : **Item\_count**

Group/Color : **name[item]**



1. X축의 데이터 범위가 너무 넓어서 필터에서 값을 일부 제한합니다. 다음처럼 값을 선택하고 Apply 버튼을 누르면 위와 같은 그래프를 볼 수 있습니다.

가격이 비쌀수록 판매 수량도 적어지는 추이가 관찰 됩니다.

