



Data Analysis



# Amazon Business Research

**Date:** 01/11/2024

**Time:** 08:00 AM

**Venue:** Thanh Pho Ho Chi Minh

## Concept of the Project

Dự án "**Amazon Business Research**" tập trung vào việc phân tích các yếu tố ảnh hưởng đến hiệu suất giao hàng và mức độ hài lòng của khách hàng. Dữ liệu được sử dụng trong dự án có nguồn từ Kaggle, bao gồm thông tin về đơn hàng, người giao hàng, tình trạng giao thông, thời tiết, và các chi tiết khác liên quan đến quá trình giao hàng.

### Taks

**Nhiệm vụ của dự án** bao gồm: phân tích yêu cầu khách hàng để xác định các yếu tố ảnh hưởng đến thời gian giao hàng; làm sạch và chuẩn bị dữ liệu bằng SQL Server; xử lý và chuẩn hóa dữ liệu với Power Query; mô hình hóa dữ liệu theo Star Schema và thiết lập mối quan hệ giữa các bảng; phân tích dữ liệu bằng DAX trong Power BI để tạo các chỉ số và cột tính toán; xây dựng báo cáo và dashboard để trực quan hóa dữ liệu; và cuối cùng là đưa ra kết luận và đề xuất các giải pháp tối ưu hóa quy trình giao hàng.

### Source Data

Nguồn: Được cung cấp miễn phí bởi người dùng VikramXD.

Link: <https://www.kaggle.com/datasets/vikramxd/amazon-business->

research-analyst-dataset.

## Tools

SQL Sever

Power BI Desktop

## Transformation

Power BI Desktop

## Visualization

Power BI Desktop

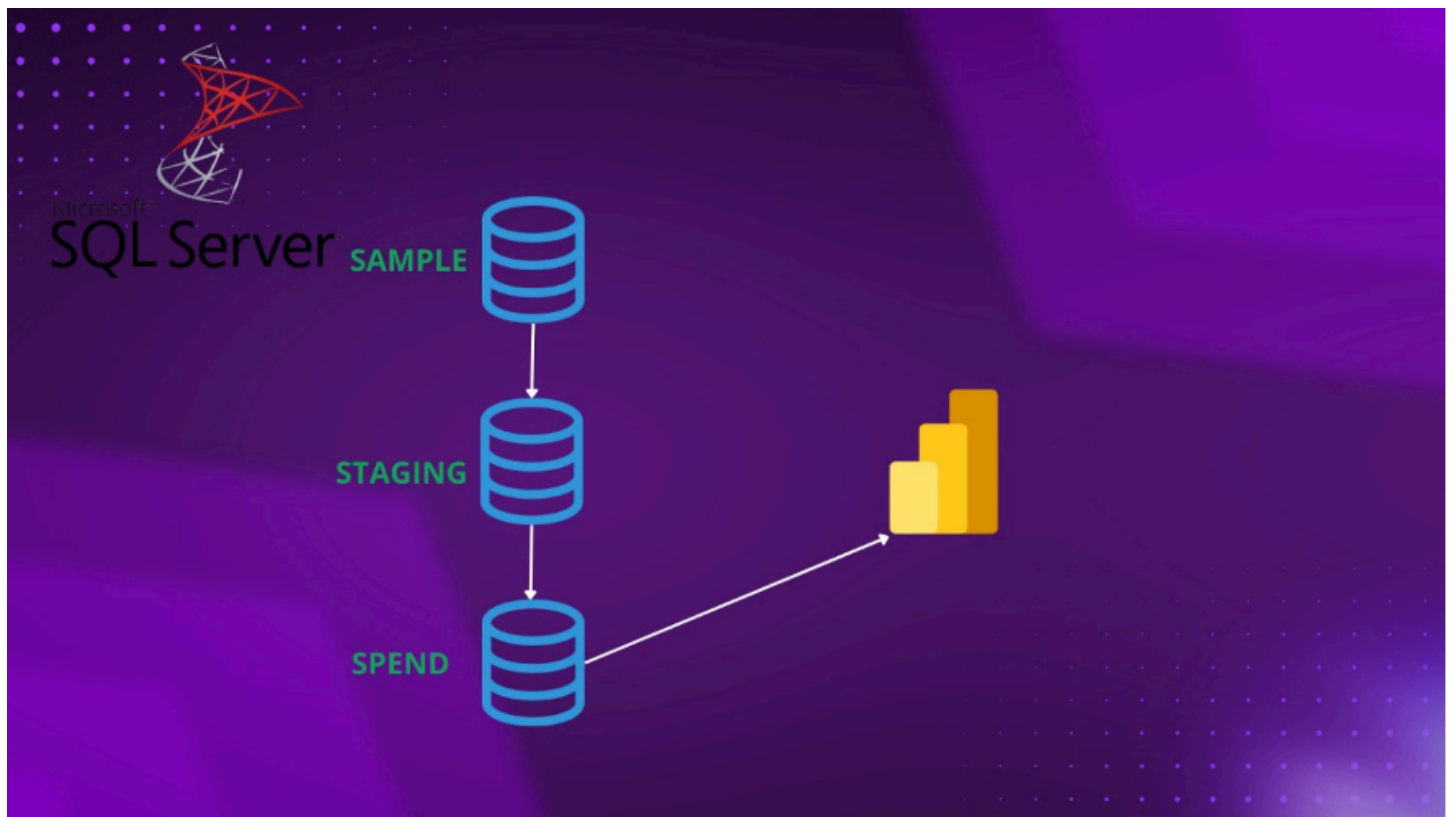
## Người thực hiện dự án

**Nguyen Thanh Dat**  
Manager

---

# Solution Architecture Overview

## Management and storage



Trong phạm vi dự án này, SQL được sử dụng để lưu trữ - xử lý:

**SPAMLE:** Database nơi chứa dữ liệu gốc.

**Staging:** Database, nơi chứa dữ liệu xử lý.

**Spend:** Nơi chứa dữ liệu hoàn chỉnh.

## Data Storytelling

### ĐẶT VẤN ĐỀ:

**Thực trạng hiện nay:** Nhiều yếu tố ảnh hưởng đến thời gian giao hàng, tuy nhiên mức độ ảnh hưởng của từng yếu tố lên hiệu suất chung và mức độ hài lòng của khách hàng vẫn chưa rõ ràng.

**Mục tiêu:** Truyền tải các thông tin từ dữ liệu để tối ưu hóa quy trình giao hàng và hiểu rõ mức độ hài lòng của khách hàng

### XÁC ĐỊNH CÂU CHUYỆN:

**Giải thích cốt lõi:** Giải thích yếu tố nào ảnh hưởng nhiều nhất đến thời gian giao hàng, nhằm xác định cơ hội cải thiện quy trình.

**Mục tiêu cụ thể:** Nâng cao hiệu suất giao hàng và mức độ hài lòng của khách hàng bằng cách làm nổi bật các mối tương quan và xu hướng trong dữ liệu giao hàng.

**Đề xuất giải pháp:** Đề xuất các cải tiến dựa trên dữ liệu, chẳng hạn như tối ưu hóa lộ trình dựa trên dữ liệu giao thông hoặc điều chỉnh nhân lực dựa trên điều kiện thời tiết.

## LÀM SẠCH DỮ LIỆU

### SQL Sever:

**Các vấn đề thường gặp:** Giá trị bị thiếu (NULL), dữ liệu không nhất quán, giá trị ngoại lệ (outliers) và dữ liệu không hợp lệ.

**Vấn đề trong dự án:** Trong bộ dữ liệu này, xuất hiện giá trị bị thiếu, NaN, không đúng kiểu dữ liệu.

### Link xử lý:

[🌐 Amazon-Business-Research/SQL Sever/query.sql at main · datntk123...](#)

## Mô hình hóa dữ liệu

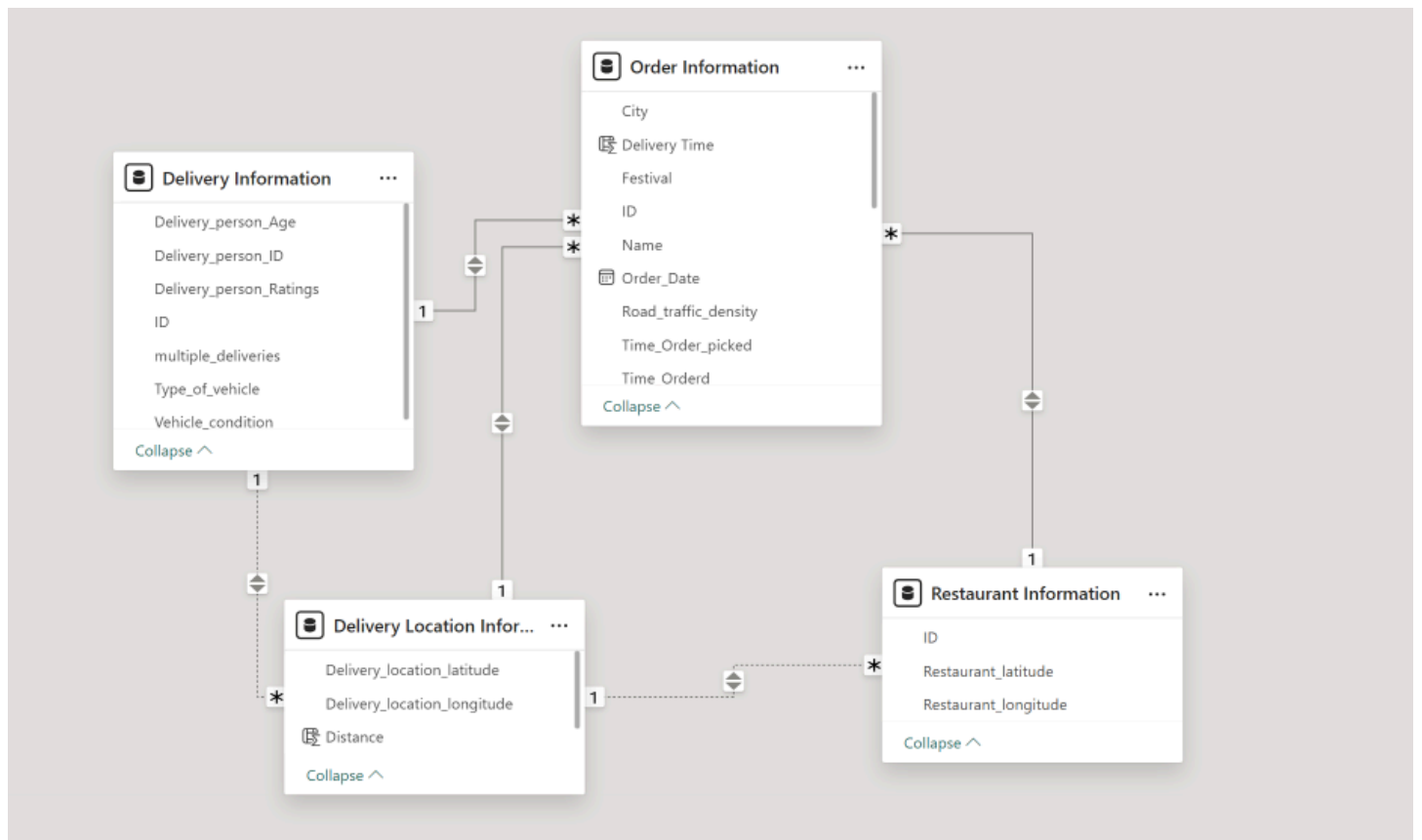
### Power BI:

**Bảng Delivery Information** lưu trữ thông tin về người giao hàng và tình trạng giao hàng, hỗ trợ phân tích hiệu suất giao hàng.

**Bảng Restaurant Information** và **Delivery Location Information** cung cấp thông tin về vị trí nhà hàng và địa điểm giao hàng, giúp tính toán quãng đường và thời gian giao hàng.

**Bảng Order Information** lưu trữ chi tiết về đơn hàng, thời gian đặt và giao hàng, cùng các yếu tố môi trường như thời tiết và mật độ giao thông, từ đó hỗ trợ việc phân tích các yếu tố ảnh hưởng đến thời gian giao hàng.

### RELATIONSHIPS:



**Delivery Information** liên kết với **Delivery Location Information** qua cột **ID** (1-nhiều). Điều này cho phép mỗi thông tin giao hàng liên kết với một vị trí giao hàng cụ thể.

**Order Information** liên kết với **Restaurant Information** qua cột **ID** (1-nhiều). Mỗi đơn hàng liên kết với một nhà hàng cụ thể.

**Order Information** cũng liên kết với **Delivery Location Information** qua cột **ID** (1-nhiều). Điều này giúp mỗi đơn hàng liên kết với một địa điểm giao hàng.

**Delivery Information** liên kết với **Order Information** qua cột **ID** (1-nhiều). Điều này cho phép theo dõi thông tin người giao hàng cho từng đơn hàng cụ thể.

**Xử lý dữ liệu DAX:**

BẢNG THÔNG TIN MEASURE CHI TIẾT		
Tên measure	Mục tiêu	Ghi chú
Delivery Time	Tính khoảng thời gian từ khi đơn hàng được đặt đến khi được lấy.	Thay thế giá trị rỗng (NULL) bằng 0 và loại bỏ các giá trị âm.
Average Delivery Time by Vehicle Condition	Phân tích xem tình trạng phương tiện có ảnh hưởng đến thời gian giao hàng không.	Dữ liệu được nhóm theo tình trạng phương tiện từ bảng Delivery Information.
Average Delivery Time by Weather	Xác định mức độ ảnh hưởng của thời tiết đến thời gian giao hàng.	Sử dụng cột Weather để phân tích từng điều kiện thời tiết.
Average Delivery Time by Traffic	Phân tích mức độ ảnh hưởng của mật độ giao thông đến thời gian giao hàng.	Sử dụng cột Road_traffic_density để phân tích từng mức độ giao thông.
Average Delivery Time by Weather and Traffic	Kết hợp thời tiết và mật độ giao thông để phân tích tổng hợp.	Phân tích giao thoa giữa Weather và Road Traffic Density.
Total Deliveries by Weather and Traffic	Phân tích số lượng giao hàng dựa trên thời tiết và mật độ giao thông.	Tính tổng số giao hàng từ cột multiple_deliveries trong điều kiện giao thông và thời tiết.
Total_Daily_Orders	Tính tổng số đơn hàng theo ngày thường.	Lọc dữ liệu với điều kiện Festival = 'NO'.
Total_Holiday_Orders	Tính tổng số đơn hàng trong dịp lễ.	Lọc dữ liệu với điều kiện Festival = 'YES'.
Avg Delivery Time by City	Phân tích thời gian giao hàng trung bình theo từng thành phố.	Sử dụng bộ lọc City để phân tích hiệu suất từng khu vực.
Avg Ratings by City	Tính điểm trung bình của người giao hàng theo thành phố.	Loại bỏ giá trị NULL trong cột Delivery_person_Ratings và nhóm theo City.

Link Công thức Dax:

🌐 [Amazon-Business-Research/Power BI Dax at main · datntk1234k/Am...](#)

Dashboard

Phần trực quan hóa của báo cáo gồm **4 trang (pages)**, mỗi trang tập trung vào các khía cạnh quan trọng của hiệu suất giao hàng:

Page 1: Vehicle Condition

Phân tích ảnh hưởng của tình trạng xe (kém, trung bình, tốt) đến số lượng đơn hàng và thời gian giao hàng.

Page 2: Weather and Traffic

Đánh giá tác động của thời tiết và mật độ giao thông đến thời gian giao hàng và hiệu suất vận chuyển.

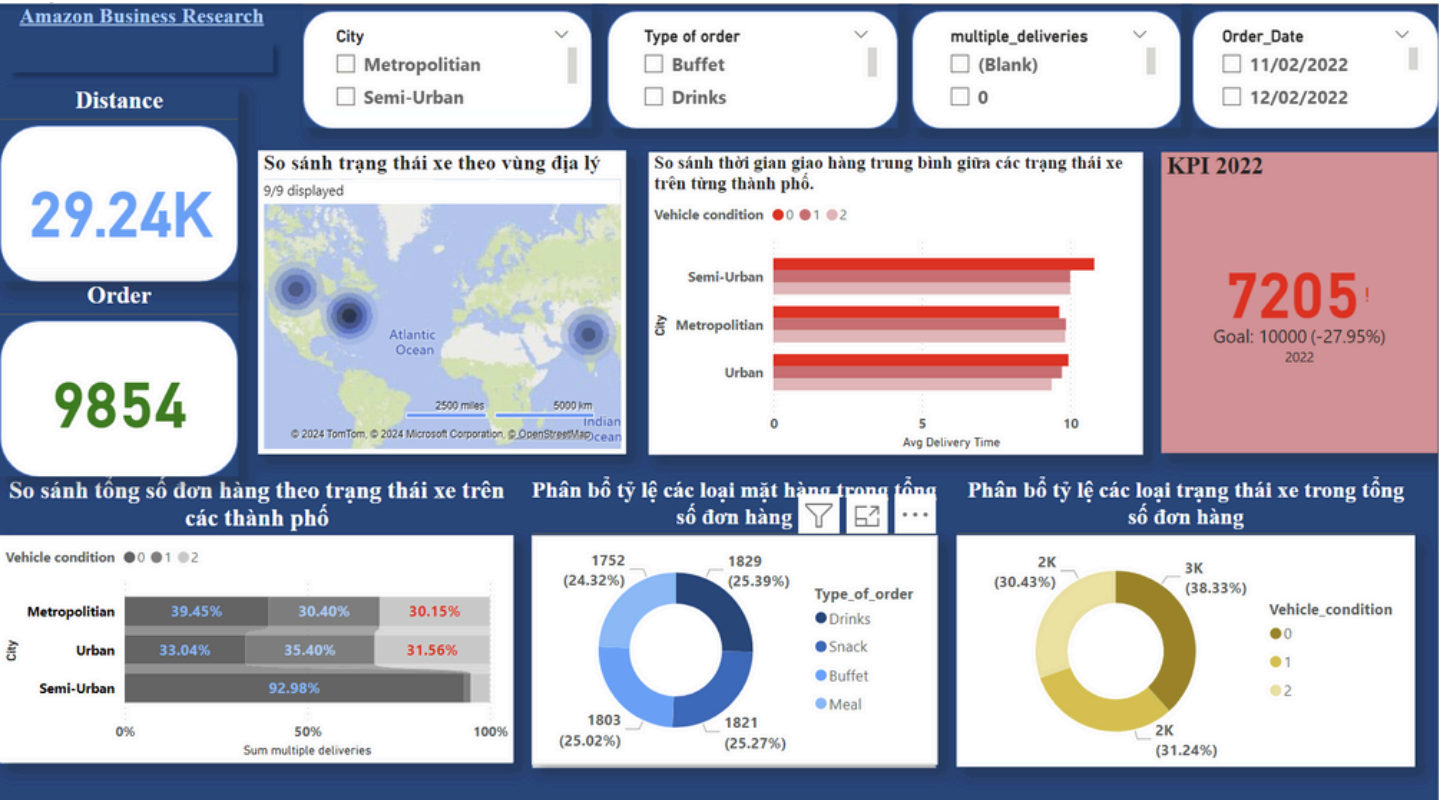
Page 3: Festival City

So sánh số lượng đơn hàng và thời gian giao hàng giữa ngày thường và ngày lễ tại các khu vực khác nhau.

Page 4: Detail

Tổng hợp các yếu tố chi tiết như độ tuổi người giao hàng, số đơn hàng giao cùng lúc, và thời gian giao hàng trung bình theo thành phố.

Page 1: Vehicle Condition



Biểu đồ: So sánh tổng số đơn hàng theo tình trạng xe ở từng thành phố



**Semi-Urban:** Xe trạng thái kém (0) chiếm tới **92.98%** số lượng đơn hàng, phản ánh tần suất sử dụng xe cao và sự phụ thuộc lớn vào xe kém trong khu vực này.

**Urban:** Xe trung bình (1) dẫn đầu với **35.40%**, nhưng xe kém (0) cũng chiếm tỷ lệ lớn (**33.04%**).

**Metropolitan:** Xe kém (0) tiếp tục đứng đầu với **39.45%**, cho thấy tình trạng xe không quá khác biệt với các khu vực khác.

### **Biểu đồ: Phân bố tỷ lệ các loại mặt hàng trong tổng số đơn hàng**

Các loại mặt hàng được đặt có tỷ lệ gần như cân bằng:

**Drinks:** Chiếm tỷ lệ cao nhất (**25.39%**).

**Snack:** Gần bằng Drinks (**25.27%**).

**Buffet và Meal:** Cùng chiếm khoảng **25%**.

Sự khác biệt nhỏ này cho thấy khách hàng có nhu cầu đa dạng về loại sản phẩm.

### **Biểu đồ: Phân tích tỷ lệ các loại trạng thái xe trong tổng số đơn hàng**

Xe trạng thái kém (0) chiếm tỷ lệ cao nhất (**38.33%**), tiếp theo là xe trung bình (1) với **31.24%** và xe tốt (2) chiếm **30.43%**.

Mặc dù chênh lệch giữa các loại xe không lớn, xe kém vẫn đóng vai trò chính trong hệ thống giao hàng.

### **Biểu đồ: So sánh thời gian giao hàng trung bình giữa các trạng thái xe theo thành phố**

**Semi-Urban:** Xe kém có thời gian giao hàng trung bình dài nhất (**hơn 10 phút**), trong khi xe trung bình và tốt giao nhanh hơn đáng kể.

**Urban:** Xe kém vẫn có thời gian giao hàng dài hơn các loại xe khác, nhưng mức chênh lệch nhỏ.



**Metropolitan:** Xe kém có thời gian giao hàng trung bình thấp nhất, hiệu quả hơn so với xe trung bình và xe tốt.

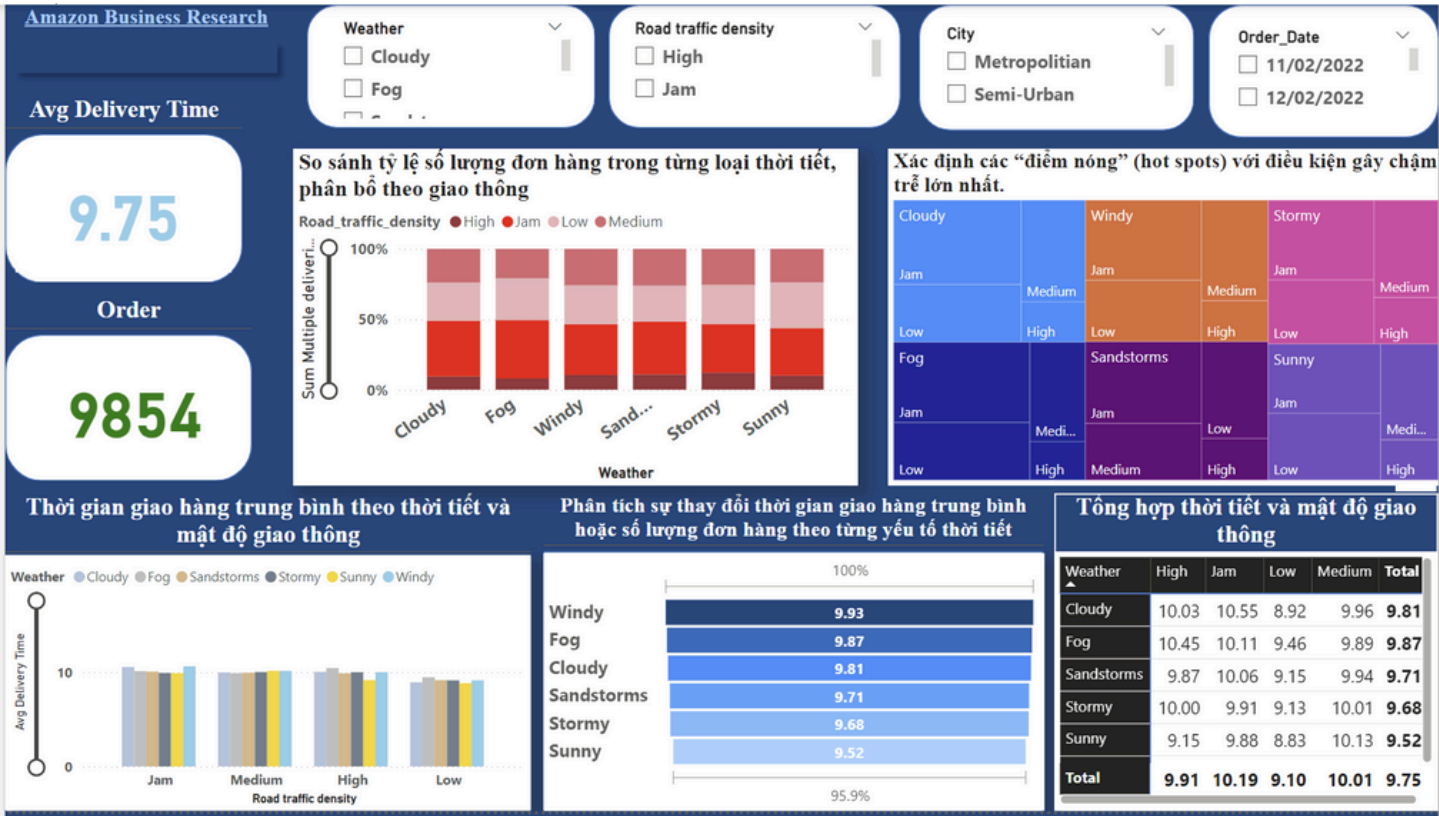
**Biểu đồ: KPI (Mục tiêu và thực tế)**

**KPI mục tiêu: 10,000 đơn hàng.**

**Thực tế đạt được: 7,205 đơn hàng, thiếu hụt 2,795 đơn hàng.**

Nguyên nhân tiềm ẩn có thể đến từ việc sử dụng xe kém hiệu quả, phân bổ nguồn lực chưa hợp lý, hoặc các chiến lược vận hành chưa tối ưu.

**Page 2: Weather and Traffic**



**Biểu đồ: So sánh tỷ lệ số lượng đơn hàng trong từng loại thời tiết, phân bổ theo giao thông**

**Tổng số đơn hàng:** Số lượng đơn hàng trong các loại thời tiết (Cloudy, Fog, Windy, Sandstorm, Stormy, Sunny) gần như tương đồng, không có sự chênh lệch đáng kể.

**Phân bổ theo giao thông:**

**Low traffic:** Chiếm tỷ lệ lớn nhất trong tất cả các điều kiện thời tiết, cho thấy giao hàng thuận lợi chủ yếu trong điều kiện giao thông thông thoáng.

**Medium và Jam traffic:** Tăng rõ rệt trong điều kiện thời tiết xấu như Stormy và Fog.

**Biểu đồ: Xác định các điểm nóng với điều kiện gây chậm trễ lớn nhất**

**Jam traffic** (Tắc nghẽn giao thông): Là yếu tố chính gây chậm trễ, đặc biệt trong các điều kiện thời tiết như Stormy, Fog, và Windy.

**Medium traffic:** Có tác động trung bình, thường xuất hiện ở các khu vực có mật độ giao thông không quá đông đúc.

**Low traffic:** Thời gian giao hàng nhanh hơn rõ rệt, chiếm ưu thế trong các điều kiện như Sunny và Sandstorm.

**High traffic:** Chiếm tỷ lệ nhỏ nhất

**Biểu đồ: Thời gian giao hàng trung bình theo thời tiết và mật độ giao thông**

**Jam:** Thời gian giao hàng cao nhất ở tất cả các loại thời tiết.

**Medium và High:** Có thời gian giao hàng trung bình giảm nhẹ so với Jam.

**Low traffic density:** Thời gian giao hàng ngắn nhất và ổn định hơn.

**Biểu đồ: “Điểm nóng” (Hotspots) gây chậm trễ lớn nhất**

**Windy và Jam:** Nổi bật với tỷ lệ lớn gây chậm trễ.

**Stormy và Medium** cũng xuất hiện với tần suất cao.

**Fog và Jam:** Là một điểm “nóng” đáng chú ý khác.

Các điều kiện **Sunny và Sandstorms** kết hợp với Jam và Medium cũng đóng góp vào thời gian giao hàng chậm.

**Biểu đồ: Phân tích thời gian giao hàng trung bình**

**Windy** có thời gian giao hàng trung bình cao nhất (9.93), tiếp theo là:

**Fog:** 9.87

**Cloudy:** 9.81

**Sandstorms:** 9.71

**Stormy:** 9.68

**Sunny:** 9.52 (thấp nhất).

Như vậy, thời tiết **gió mạnh (Windy)** và **sương mù (Fog)** ảnh hưởng đáng kể đến thời gian giao hàng.

### **Bảng tổng hợp thời tiết và mật độ giao thông**

**Điều kiện Jam (ùn tắc giao thông)** có thời gian giao hàng trung bình cao nhất (10.19).

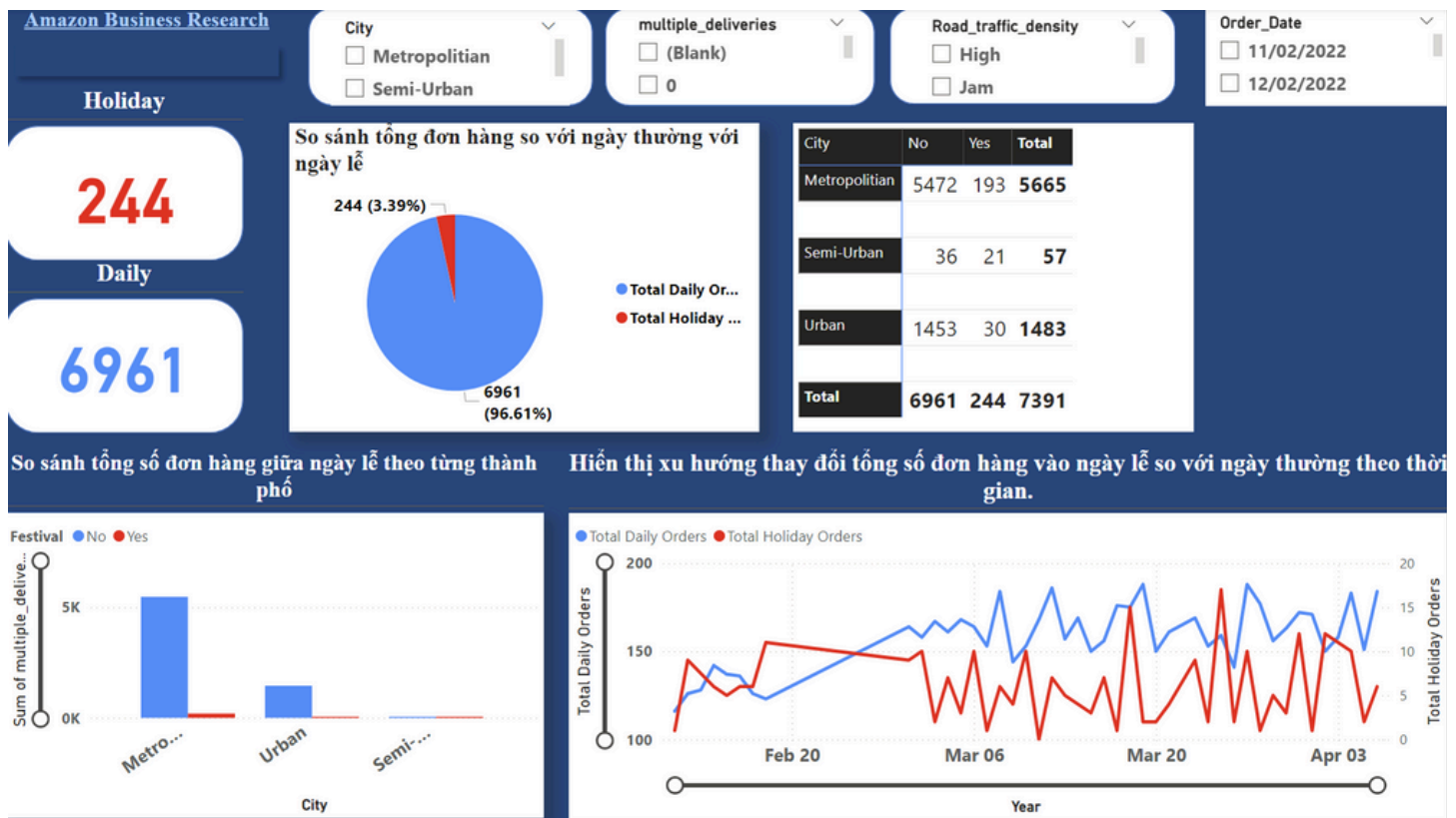
**Low (giao thông thông thoáng)** có thời gian giao hàng thấp nhất (9.10).

Tổng hợp thời tiết cho thấy:

**Fog và Jam** có thời gian giao hàng cao nhất (10.45).

**Sunny và Low** có thời gian giao hàng thấp nhất (9.15).

**Page 3: Festival City**



## Biểu đồ: So sánh tổng số đơn hàng vào ngày lễ với ngày thường theo từng thành phố

### Metro City:

Ngày thường có tổng số đơn hàng rất cao, áp đảo so với ngày lễ, cho thấy nhu cầu không phụ thuộc nhiều vào các dịp lễ.

### Urban City:

Ngày thường vẫn chiếm ưu thế, nhưng mức chênh lệch so với ngày lễ nhỏ hơn so với Metro.

### Semi-Urban City:

Tổng số đơn hàng ở cả hai loại ngày đều rất thấp, gần như không đáng kể, cho thấy mức độ tiêu thụ yếu ở khu vực này.

## Biểu đồ: Hiển thị xu hướng thay đổi tổng số đơn hàng vào ngày lễ so với ngày thường theo thời gian

### Xu hướng ngày thường (xanh nhạt):

Tổng số đơn hàng ngày thường ổn định hơn, dao động trong khoảng từ 100 đến 200 đơn hàng.

Tăng trưởng đều đặn, với một vài thời điểm giảm nhẹ nhưng nhanh chóng phục hồi.

### **Xu hướng ngày lễ (xanh đậm):**

Số đơn hàng vào ngày lễ thấp hơn rất nhiều so với ngày thường, chỉ dao động từ 0 đến 20 đơn hàng.

Có các đợt tăng vọt (spikes) vào một số ngày lễ nhất định nhưng không duy trì lâu dài, thể hiện tính thời điểm.

### **Biểu đồ: So sánh tổng số đơn hàng so với ngày thường với ngày lễ**

#### **Tỷ lệ đơn hàng:**

Ngày thường (Total Daily Orders - 96.61%): Áp đảo với tổng số 6961 đơn hàng, chiếm phần lớn lượng đơn hàng.

Ngày lễ (Total Holiday Orders - 3.39%): Chỉ đạt 244 đơn hàng, chiếm một tỷ lệ rất nhỏ trong tổng số.

#### **Tương quan ngày thường và ngày lễ:**

Đơn hàng ngày thường cao gấp nhiều lần ngày lễ, cho thấy nhu cầu mua sắm không tập trung vào các dịp lễ.

### **Page 4: Detail**



## Biểu đồ: Hiện thị thời gian giao hàng trung bình theo thành phố

Semi-Urban có thời gian giao hàng trung bình cao nhất, gần chạm mức 10.

Metropolitan có thời gian giao hàng trung bình xếp thứ hai, nằm trong khoảng từ 7 đến 8.

Urban có thời gian giao hàng trung bình thấp nhất, khoảng dưới 7.

## Biểu đồ: Số lượng đơn hàng giao cùng lúc có ảnh hưởng đến thời gian giao hàng

Khi số lượng đơn hàng giao cùng lúc (Multiple Deliveries) tăng lên, thời gian giao hàng trung bình (Avg Delivery Time) có xu hướng tăng:

**0 đơn hàng:** Thời gian giao hàng trung bình gần mức 9.6 - 9.8 (xuýt soát so với 1 đơn hàng).

**1 đơn hàng:** Thời gian giao hàng trung bình khoảng 9.8.

**2 đơn hàng:** Thời gian giao hàng trung bình tăng lên gần 10.

**3 đơn hàng:** Có xu hướng giảm nhẹ so với khi giao 2 đơn hàng.

**Biểu đồ: Điểm đánh giá đơn hàng trung bình theo thành phố**

**Urban:**

Có điểm đánh giá cao nhất, cho thấy dịch vụ giao hàng tốt nhờ cơ sở hạ tầng phát triển và giao thông hiệu quả.

**Metropolitan:**

Điểm đánh giá thấp hơn Urban, cho thấy chất lượng dịch vụ chưa đồng đều, ảnh hưởng bởi tắc nghẽn giao thông và mật độ dân số cao.

**Semi-Urban:**

Có điểm đánh giá thấp nhất, phản ánh hạn chế về cơ sở hạ tầng và chất lượng dịch vụ không ổn định ở khu vực bán đô thị.

**Vấn đề:**

**Ảnh hưởng của tình trạng xe đến hiệu suất giao hàng:**

Xe trong trạng thái kém ảnh hưởng nghiêm trọng đến số lượng đơn hàng và thời gian giao hàng.

Mức độ sử dụng xe kém cao ở các khu vực bán đô thị và thành phố lớn.

**Ảnh hưởng của thời tiết và mật độ giao thông:**

Tình trạng tắc nghẽn giao thông, đặc biệt là trong các điều kiện thời tiết xấu (gió mạnh, sương mù), dẫn đến thời gian giao hàng cao.

Mức độ giao thông cao làm giảm hiệu suất giao hàng, đặc biệt là trong thời gian bão, sương mù, và gió mạnh.

**Ảnh hưởng của các dịp lễ:**

Lượng đơn hàng vào ngày lễ thấp hơn rất nhiều so với ngày thường, điều này có thể làm giảm hiệu quả vận hành trong các kỳ lễ.

**Tác động của số lượng đơn hàng giao cùng lúc:**



Khi số lượng đơn hàng giao cùng lúc tăng, thời gian giao hàng trung bình cũng tăng, ảnh hưởng đến hiệu suất giao hàng.

## **Giải pháp:**

### **Tối ưu hóa việc sử dụng xe:**

Đầu tư vào bảo trì xe và thay thế xe kém chất lượng, đặc biệt là ở các khu vực bán đô thị, để nâng cao hiệu suất giao hàng.

Cải thiện việc phân bổ xe sao cho các xe có trạng thái tốt được phân bổ cho các khu vực có nhu cầu cao.

### **Điều chỉnh quy trình vận hành dựa trên thời tiết và mật độ giao thông:**

Tối ưu hóa lộ trình giao hàng dựa trên dữ liệu giao thông thời gian thực và dự báo thời tiết, giúp giảm thiểu thời gian giao hàng trong điều kiện thời tiết xấu.

Đưa ra các chiến lược giao hàng linh hoạt hơn vào những ngày có mật độ giao thông cao.

### **Tăng cường marketing và giảm chi phí trong dịp lễ:**

Đẩy mạnh các chiến lược marketing để tăng đơn hàng vào các dịp lễ, giúp cân bằng lượng đơn hàng giữa ngày lễ và ngày thường.

Cải thiện dịch vụ và ưu đãi vào các ngày lễ để thu hút khách hàng.

### **Tăng cường đào tạo và tối ưu hóa quy trình giao hàng:**

Đào tạo nhân viên giao hàng để xử lý các đơn hàng một cách hiệu quả, đặc biệt khi giao nhiều đơn hàng cùng lúc.

Áp dụng các chiến lược tối ưu hóa quy trình vận hành để giảm thời gian giao hàng khi giao nhiều đơn hàng.

## **Kết thúc:**

Thông qua việc tối ưu hóa lộ trình giao hàng, cải thiện tình trạng phương tiện và ứng dụng các chiến lược linh hoạt, chúng ta có thể giải quyết hiệu

quả những thách thức trong quy trình giao hàng, nâng cao chất lượng dịch vụ và đáp ứng kỳ vọng của khách hàng.

