**BÁO CÁO PHÂN TÍCH KHẢ NĂNG THÀNH CÔNG CỦA DỰ ÁN BIKE-SHARE CYCLISTIC**

1. **TỔNG QUAN**

Báo cáo phân tích này nhằm phân tích mức độ thành công của dự án Bike-share Cyclistic tại thời điểm tháng 3 năm 2023 thông qua số lượng thành viên đăng ký dịch vụ của Cyclistic.

* Phương pháp và công cụ: Phân tích sữ được thực hiện bằng các công cụ phân tích dữ liệu khác nhau, bao gồm: Excel, Power BI và Python. Dữ liệu được phân tích bao gồm dữ liệu thuê xe của khách hàng trong quý 1 năm 2023.
* Mô tả dữ liệu: Dữ liệu trên được thu thập thông qua quá trình thuê xe của khách hàng, bao gồm các thông tin sau:

|  |  |  |
| --- | --- | --- |
| **STT** | **Thuộc tính** | **Mô tả** |
| 1 | ride\_id | Mã thuê xe |
| 2 | rideable\_type | Loại xe thue |
| 3 | started\_at | Thời gian bắt đầu thuê xe |
| 4 | ended\_at | Thời gian trả xe |
| 5 | start\_station\_name | Tên trạm thuê xe |
| 6 | start\_station\_id | Mã trạm thuê xe |
| 7 | end\_station\_name | Tên trạm trả xe |
| 8 | end\_station\_id | Mã trạm trả xe |
| 9 | start\_lat | Vĩ độ của trạm thuê xe |
| 10 | start\_lng | Kinh độ của trạm thuê xe |
| 11 | end\_lat | Vĩ độ của trạm trả xe |
| 12 | end\_lng | Kinh độ của trạm trả xe |
| 13 | member\_casual | Loại khách hàng thuê xe |

* Câu hỏi nghiên cứu:

+ Thành viên và người đi xe bình thường sử dụng xe đạp khác nhau như thế nào?

+ Tại sao những người đi xe bình thường lại mua tư cách thành viên hàng năm của Cyclistic?

+ Làm cách nào để Cyclistic có thể tác động đến những người đi xe bình thường trở thành thành viên?

1. **QUY TRÌNH XỬ LÝ DỮ LIỆU**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Xử lý** | **Mô tả chi tiết** | **Công cụ** | **Thời gian** |
| 1 | Tổng hợp dữ liệu | Tổng hợp dữ liệu thuê xe trong tháng 1, 2, 3 thành dữ liệu thuê xe quý 1 (Qtr1-divvy-tripdata) | MS Excel | 13:10 24/04/2023 |
| 2 | Sắp xếp dữ liệu | Sắp xếp dữ liệu theo ngày thuê xe từ cũ nhất đến mới nhất | MS Power BI | 19:47 24/04/2023 |
| 3 | Lọc dữ liệu | Kiểm tra các dữ liệu bị null ở cột start\_station\_name, start\_station\_id, end\_station\_name, end\_station\_id xem có ảnh hưởng đến quá trình phân tích hay không | MS Power BI | 20:06 24/04/2023 |
| 4 | Sửa dữ liệu | Chỉnh sửa dữ liệu bị lỗi của cột member\_casual | MS Excel | 20:10 24/04/2023 |
| 5 | Tạo cột mới | Sử dụng dax để tạo cột duration là khoảng thời gian từ lúc thuê xe đến lúc trả xe, đặt data format là hh:mm:ss | MS Power BI | 20:15 24/04/2023 |
| 6 | Tạo cột mới | Sử dụng dax để tạo cột weekday là ngày thuê xe trong tuần (Mon, Tue,…) | MS Power BI | 20:27 24/04/2023 |
| 7 | Sửa dữ liệu | Đồng nhất kiểu dữ liệu của cột start\_station\_id và end\_station\_id (chuyển về kiểu text) | MS Excel | 21:02 24/04/2023 |
| 8 | Xóa dữ liệu | Xóa dữ liệu null của cột start\_station\_id | MS Power BI | 14:30 25/04/2023 |
| 9 | Tạo hàm tính toán | Sử dụng dax để tạo hàm tính thời gian thuê xe trung bình và thời gian thuê xe lâu nhất | MS Power BI | 16:00 25/04/2023 |

**Tóm tắt dữ liệu:**

* Có 88,104 dòng dữ liệu có start\_station\_name và start\_station\_id bị null.
* Có 93,016 dòng dữ liệu có end\_station\_name và end\_station\_id bị null.
* Có 426 dòng dữ liệu có end\_lat và end\_lng bị null.
* Có 2 dòng dữ liệu bị lỗi (casuai thay vì casual) ở cột member\_casual.
* Cột start\_station\_id và end\_station\_id bị lỗi không đồng nhất kiểu dữ liệu (text và number).

1. **PHÂN TÍCH DỮ LIỆU**
2. **Tổng quan dự án quý 1 năm 2023**

Dựa vào biểu đồ phân tích, ta thấy rằng có 551,320 lượt thuê xe trong đó có 428,259 thành viên và 123,061 người thuê tự do. Có 3 phân khúc xe có thể thuê bao gồm: xe đạp cổ điển, xe đạp điện và xe đạp có trạm đỗ. Có tổng cộng 1066 trạm thực hiện việc cho thuê xe và 1028 trạm thực hiện việc nhận lại xe sau khi thuê. Dữ liệu được phân tích hầu hết ở khu vực Chicago thuộc tiểu bang Illinois nước Mỹ.

Graphical user interface, application

Description automatically generated

1. **Tình hình dự án Bike-share cyclistic**

* **Tỷ lệ thành viên trên tổng số người thuê xe**

Diagram

Description automatically generated

Tỉ lệ thành viên trong tổng số khách hàng thuê xe là 77.68% trong khi khách hàng thông thường chỉ chiếm 22.32%.

* Hầu hết mọi người thuê xe của Cyclistic đều đăng ký thành viên chứng tỏ dự án trên đang đi đúng hướng và có thể tiến hành thêm các chiến lược marketing khác để gia tăng tỉ lệ đăng ký thành viên của người dịch vụ cho thuê xe. Đồng thời tiến hành nghiên cứu tại sao những người dùng khác không đăng ký thành viên và tìm cách khắc phục.
* **Tình hình thuê xe quý 1 năm 2023**

Chart

Description automatically generated with medium confidence Chart

Description automatically generated

Có sự gia tăng số lượng lượt thuê xe(163,580 – tháng 1, 164,972 – tháng 2 và 222,768 – tháng 3) và số lượng đăng ký thành viên(130,124 – tháng 1, 128,274 – tháng 2 và 169,861 – tháng 3) trong xuyên suốt quý 1, tuy nhiên tỷ lệ thành viên trên tổng số lượt thuê xe lại có chút giảm nhẹ (79.55% – tháng 1, 77.76% – tháng 2 và 76.25% – tháng 3)

* Tuy tỉ lệ thành viên trên tổng số người thuê xe là khá cao nhưng đang có chiều hướng đi xuống. Do đó công ty cần phải duy trì số lượng thành viên cũ cũng như đề ra các giải pháp tiếp cận đối với những người chưa đăng kí thành viên sử dụng dịch vụ.

1. **Phân tích nhu cầu sử dụng dịch vụ**

* **Tình hình thuê xe trong tuần**

A picture containing line chart

Description automatically generated

Dựa vào biểu đồ ta thấy số lượng người thuê xe là khá cao đối với những ngày làm việc và thấp nhất đối với 2 ngày cuối tuần (thứ 7 và chủ nhật).

* Nhu cầu sử dụng xe đạp làm phương tiện đi lại chính (đi học, đi làm,…) khá cao. Có thể đưa ra các chiến lược giảm giá cho học sinh, sinh viên nếu đăng kí thành viên hoặc giảm giá nếu thuê xe vào cuối tuần. Càng có nhiều người sử dụng dịch vụ, số lượng thành viên sẽ ngày càng tăng.
* **Nhu cầu sử dụng dịch vụ**

**Map

Description automatically generated Chart, bar chart

Description automatically generated**

Nhu cầu thuê xe và các trạm thuê, trả xe trải dài khắp khu vực Chicago chứ không tập trung vào vào vùng nào cụ thể. Số lượng lượt thuê xe có 1 chút nhỉnh hơn cho loại xe đạp cổ điển chiếm 52.1% so với xe đạp điện 46.63%. Đối với loại xe đạp có trạm đỗ chỉ chiếm khoảng hơn 1% và toàn bộ đều được sử dụng bởi người dùng không phải là thành viên.

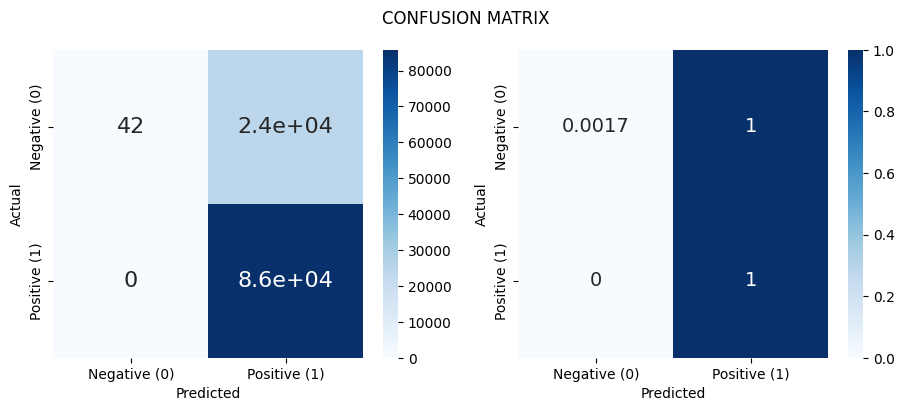
* Tỉ lệ sử dụng xe đạp cổ điển và xe đạp điện khá đồng đều, đồng thời xe đạp có trạm đỗ còn khá thấp. Qua đó cho thấy người dùng rất đa dạng về nhu cầu sử dụng chứ không quá tập trung vào loại xe nào. Quy mô của dự án cũng là khá tốt khi bao quát được toàn bộ khu vực Chicago, có thể cân nhắc mở rộng dự án nếu số lượng thành viên và nhu cầu sử dụng dịch vụ tiếp tục tăng.

1. **Phân loại thành viên và người thuê xe bình thường**
2. **Quá trình phân loại**

Phân loại theo 2 cách khác nhau và so sánh hiệu quả phân loại.

* **Cách 1: Giữ nguyên dữ liệu sau khi xử lý và tiến hành phân loại**

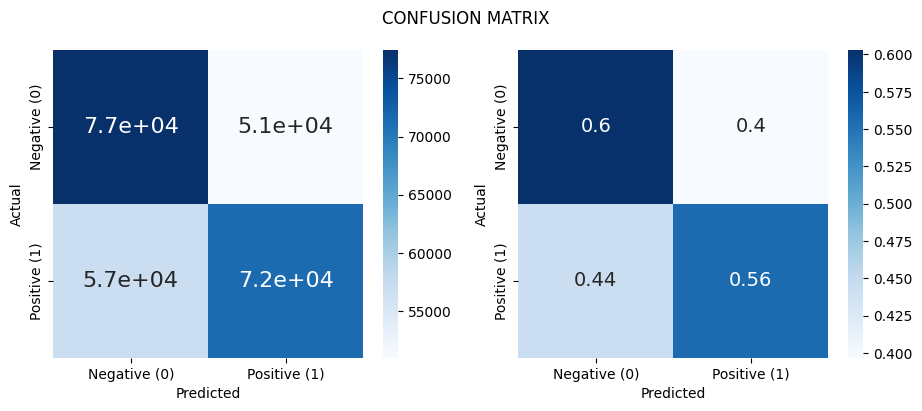
**Confusion matrix**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| Casual | 1.00 | 0.00 | 0.00 | 24448 |
| Member | 0.78 | 1.00 | 0.88 | 85790 |
| Accuracy |  |  | 0.78 | 110238 |
| Marco avg | 0.89 | 0.50 | 0.44 | 110238 |
| Weighted avg | 0.83 | 0.78 | 0.68 | 110238 |

* **Cách 2: Sử dụng kĩ thuật Smote Oversampling**

**Confusion matrix**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | Support |
| Casual | 0.58 | 0.60 | 0.59 | 128449 |
| Member | 0.58 | 0.56 | 0.57 | 128449 |
| Accuracy |  |  | 0.58 | 256898 |
| Marco avg | 0.58 | 0.58 | 0.58 | 256898 |
| Weighted avg | 0.58 | 0.58 | 0.58 | 256898 |

1. **Kết luận**

Vì dữ liệu có dấu hiệu bị thiên lệch khá lớn giữa thành viên(428,259) và người thuê bình thường(123,061) nên việc giữ nguyên dữ liệu sau xử lý để phân tích sẽ cho ra các kết quả không chính xác 100% Type 1 Error vì dữ liệu tập train về người thuê xe bình thường là quá hạn chế. Việc áp dụng kĩ thuật Over Sampling sẽ khiến dữ liệu không còn bị thiên lệch và thuận lợi hơn cho quá trình phân loại, tuy nhiên việc cân bằng dữ liệu thành viên và dữ liệu người thuê xe bình thường sẽ không đúng với thực tế, làm sai kết quả phân loại.