MEDICAL APPOINTMENT NO-SHOW ANALYSIS & RISK SCORING SYSTEM

Predictive Insights & Operational Strategy for Reducing No-Shows

EXECUTIVE SUMMARY

- 110K+ patient appointments
- Built probability models, statistical tests, and a risk scoring system
- Designed four Tableau dashboards
- Found that age, wait time, reminders, scholarship, weekday, and underlying conditions are statistically significant
- Built a risk score (Low/Medium/High) which showed high-risk patients no-show 2× more often than low-risk patients
- Goal is to find ways to decrease no-shows
- Data is from Kaggle Joni Hoppen and Aquarela Analytics

PROJECT WORKFLOW

- Data cleaning & preprocessing
- Feature engineering
- Probability analysis
- Statistical testing (Chi-Square)
- Tableau dashboards
- Risk scoring model
- Operational recommendations

DATASET OVERVIEW

- 110502 Rows representing doctor appointments from a Brazilian clinic
- Key fields include PatientID, Appointment ID, ScheduledDay, AppointmentDay, Age, SMS_received, and No-show
- Variables engineered:
  - *Days_Wait*
  - *Days_Group*
  - *Age_Group*
  - *Weekday*
  - *Condition*
  - *Aid*
  - *Reminder*
  - *Risk Score*
  - *Risk Category*

DATA CLEANING & FEATURE ENGINEERING

- Converted ScheduledDay & AppointmentDay to datetime
- Dropped rows missing ScheduledDay, AppointmentDay, or PatientID
- Filled nulls for Age with the median and Gender with Unkown
- Created Days_Wait to show the time difference between the Scheduled & Appointment Days
- Days_Wait was binned to create Days_Group
- Age was binned to create Age_Group
- Condition was made to generalize the underlying conditions
- No-show was normalized
- Same-day scheduled appointments were filtered out due to skewing the No-show data

PROBABILITY ANALYSIS

- Overall No-Show Rate: 28.5%
- Teens (36%) & Young Adult (34%) are the most at risk age group
- Seniors (21%) are the lowest at risk age group
- A 2 Month wait leads to the most No-Shows while a Week wait leads to the least
- Sending a reminder decreases the chance of a No-Show
- Being scholarship increases the chance of a No-Show
- Having a Condition decreases the chance of a No-Show except for Seniors where it increases the chance
- Wednesday has the lowest No-Show rate while Monday has the highest

CHI-SQUARE STATISTICAL SIGNIFICANCE

1

The Chi-Square test is used to examine whether two categorical variables are independent in influencing the test statistic

2

The Chi-Square test showed Gender is not significant and thus doesn't effect the probability of a patient not showing up

3

The Chi-Square test did show that the following were significant and do effect the probability: *Age Group, Weekday, Wait Time, Condition, Scholarship, & Reminder*

TABLEAU DASHBOARD – EXECUTIVE SUMMARY

- No-show KPI showing the probability of a No-Show
- Average wait KPI showing how many days the wait is
- Patient demographics of if a Reminder was sent, if they have an underlying condition, and if they receive the Scholarship
- Breaks down the number of Appointments & Patients by Risk Category
- Shows the No-Show rate of each Risk Category

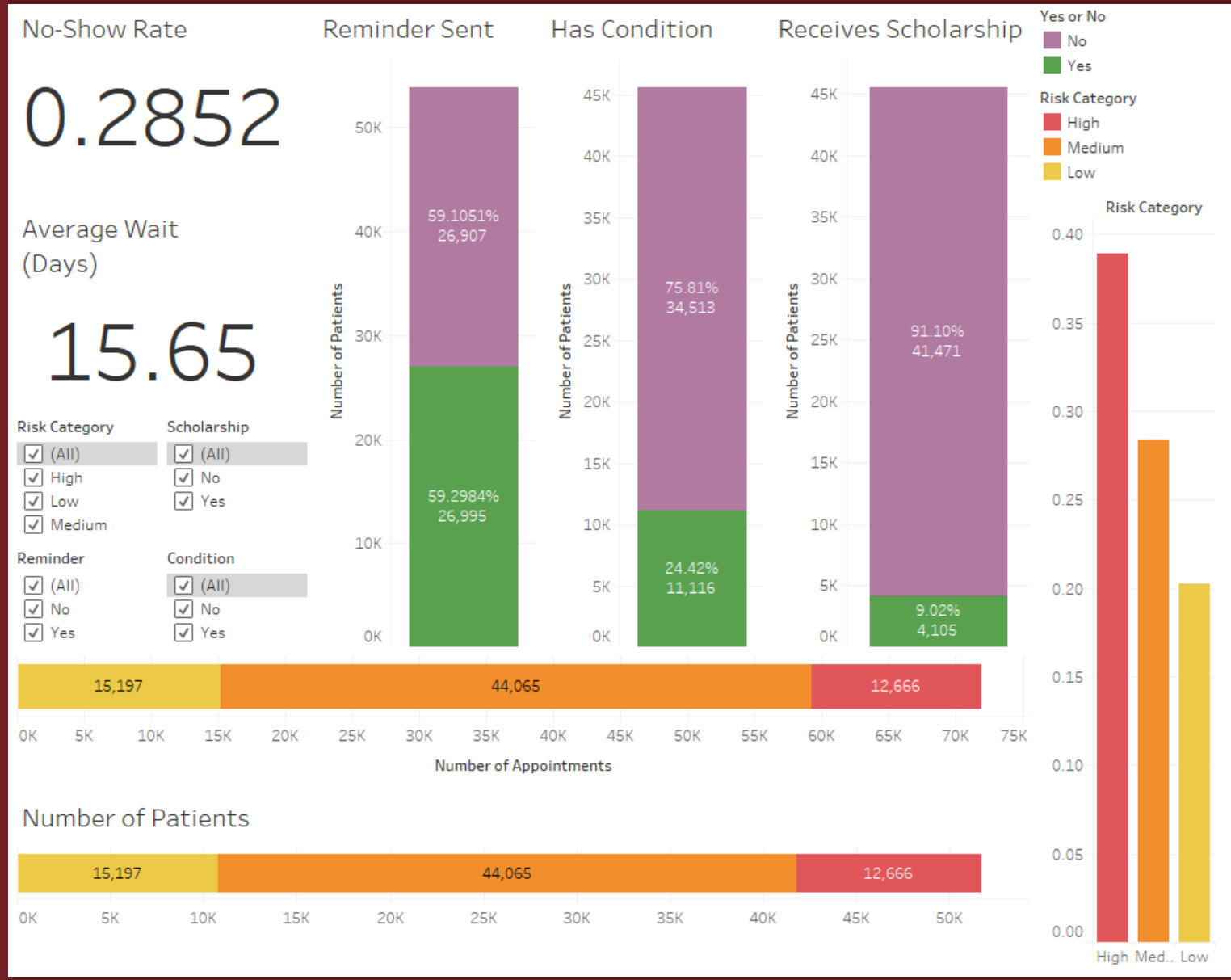


TABLEAU DASHBOARD – DRIVER ANALYSIS

- Weekday patterns showing average No-Show rate
- Age-group No-Show trends
- Wait-Time No-Show trends
- Impact of reminders, conditions, aid on No-Show Rates
- Amount of Patients in each Risk Category
- Shaded to show No-Show Rate

TABLEAU DASHBOARD – RISK SCORING

- Risk score mapping
- Category-level no-show rates
- Distribution of risk levels
- Heatmap of risk across combinations of factors

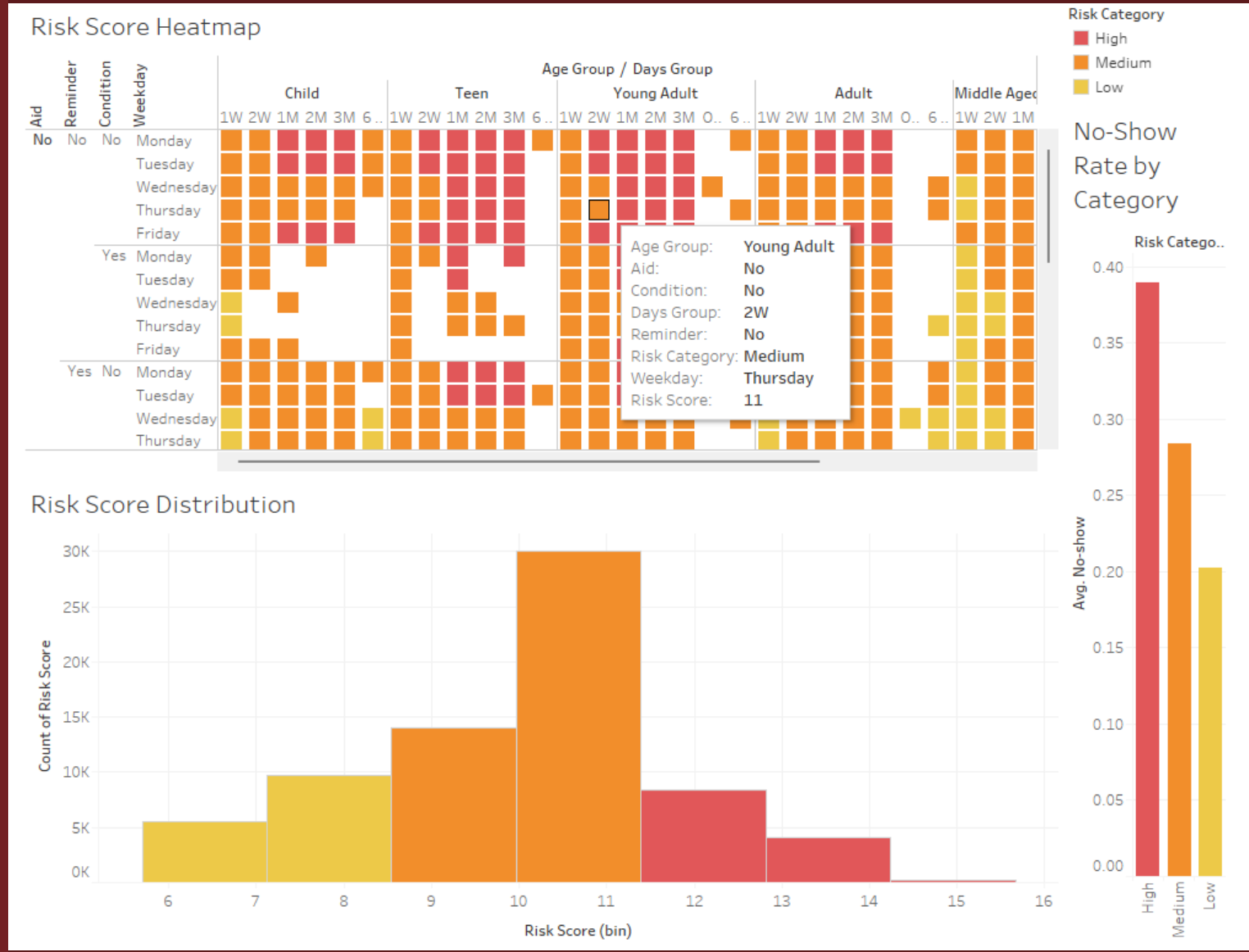
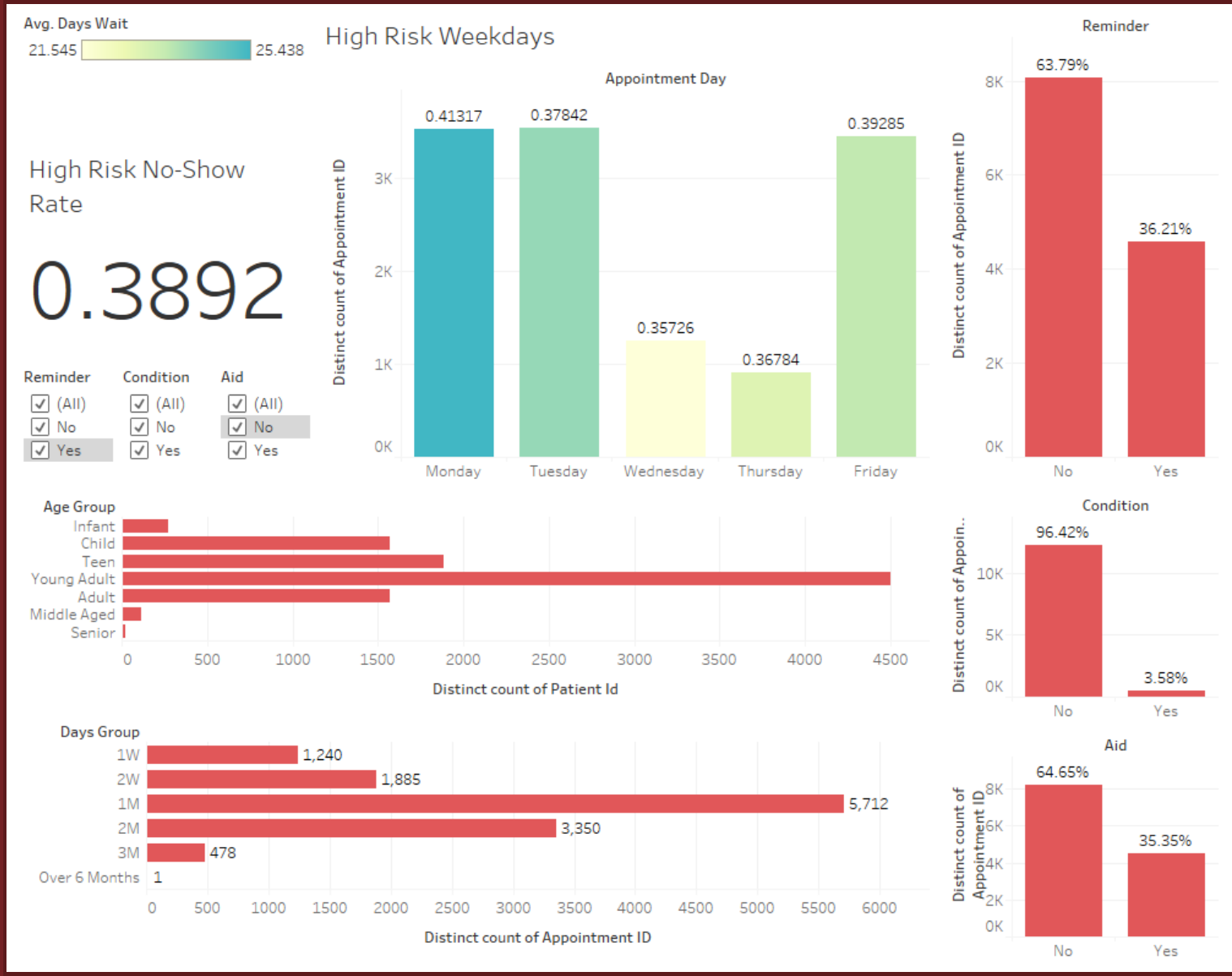


TABLEAU DASHBOARD – OPERATIONAL STRATEGY

- High Risk No-Show KPI showing the probability of a No-Show
- High-risk patient segments
- High-risk weekdays w/ average wait time
- High-risk wait times
- Impact of reminders, Condition, & Scholarship



RISK SCORING MODEL

- Points assigned based on Age group, Waiting time, Reminder, Condition, Aid, & Weekday
- Points (1-3) are assigned based on how high the probability of a no show
- 1 Points for below the average No-Show rate (less than 27%)
- 2 Points for about average (27%-33%)
- 3 Points for above average (more than 33%)
- For Age Group Senior & Middle Age received 1 Point, Adult & Child & Infant received 2 Points, while Teen & Young adult received 3 Points
- For Waiting time a Week & 6 Months & Over 6 Months received 1 Point, 2 Weeks & 3 Months received 2 Points, while 1 Month & 2 Months received 3 Points
- If a reminder was sent 1 Point was received and if no reminder was sent 2 Points were received
- If the patient has an underlying condition 1 Point was received and if not 2 Points were received
- If the patient was a Scholarship recipent 3 Points were received and if not 1 Point was received
- If the appointment was on a Thursday or Wednesday 1 Point was received and every other day received 2 Points

RISK SCORE PERFORMANCE

The Risk Score was sorted into 3 Categories: Low (Below 9), Medium (9-12), & High (Above 12)

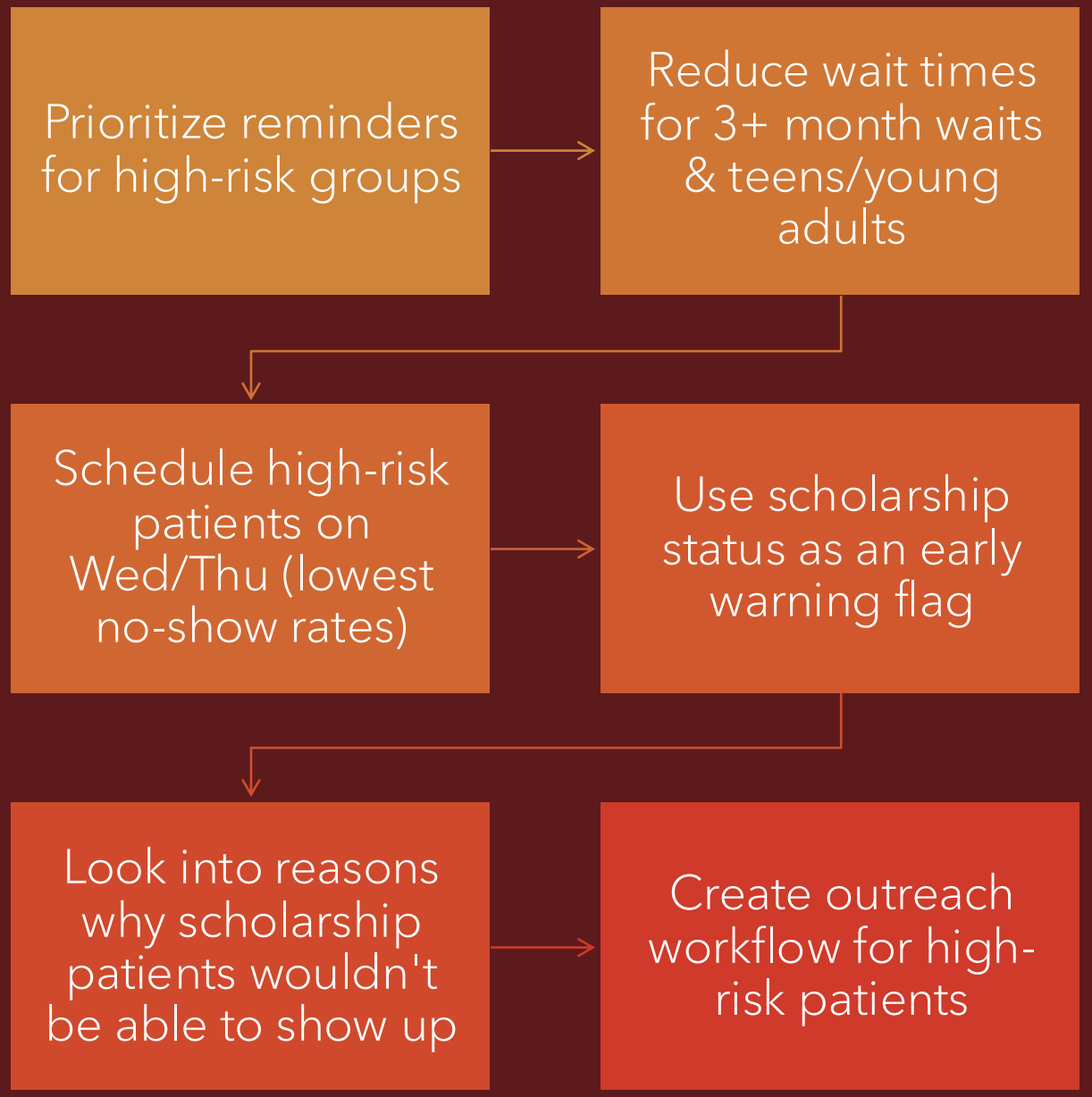
High had a No-Show Percent of 38.9%

Medium had a No-Show Percent of 28.3%

Low had a No-Show Percent of 20.2%

This was checked by checking the probability of a no-show for combinations that would result in each category

OPERATIONAL RECOMMENDATIONS



FUTURE WORK

Predictive modeling (transform risk score into logistic regression)

Intervention effectiveness simulation

Automated reminder prioritization

Look into reasons high-risk patients wouldn't show up (ex. no transportation)