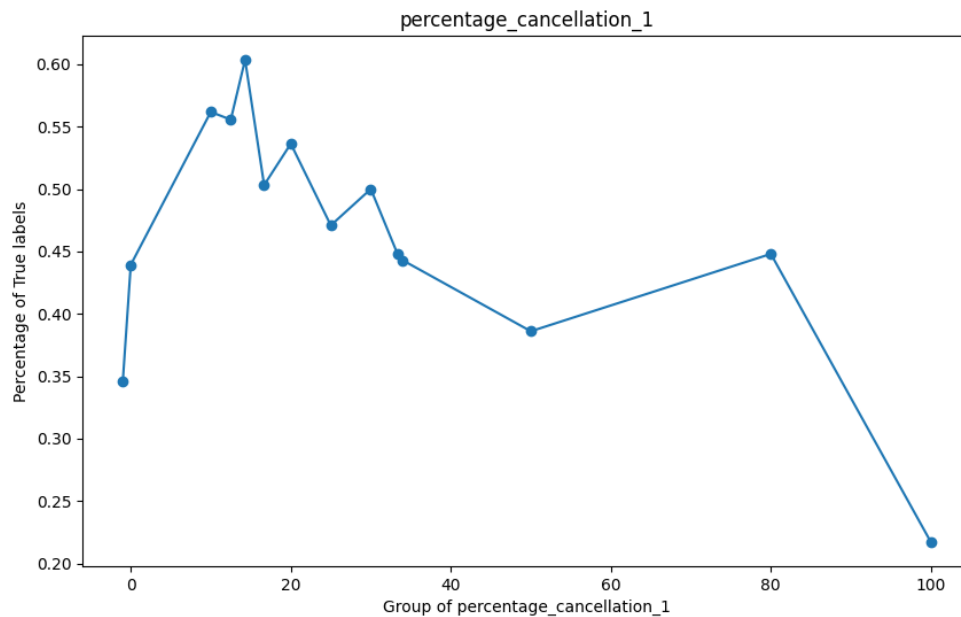
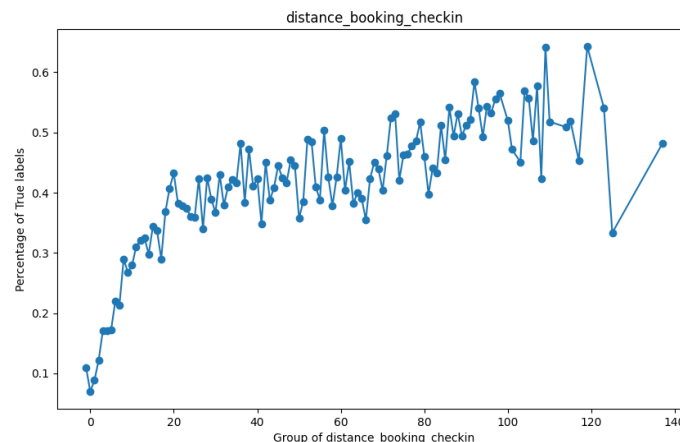


אגודה - משימה 1.2.3

הפיצ'ר הראשון שבחרנו הוא כמובן מדיניות הביטולים שסוכמה בהזמנה. פיצ'ר 'percentage_cancellation_1' מתאר את אחוז התשלום שהלקוח משלם בעת הביטול מתוך הסכום המקורי, בגרף המצורף ניתן לראות כי ככל שהאחוז גבוה יותר כך ההסתברות שהלקוח יבטל היא נמוכה יותר בצורה דרסטית (מכ%45 ומעלה היא יורדת עד לכ%20). קורולציה זו היא הגיונית משום שמדיניות המאפשרת ללקוח לבטל מבלי לשלם קנס (או קנס קטן) בוודאי תעודד את הלקוחות לבטל ונצפה שההסתברות לביטול במצב זה תהיה גבוהה יותר.



הפיצ'ר השני שאנו מניחים שיהיה משמעותי הוא המרחק בימים בין ביצוע ההזמנה לבין יום הצ'ק אין המיועד, פיצ'ר זה לא היה בדאטה המקורי אך יכולנו להרכיב אותו בקלות יחסית מהפיצ'רים שכן היו נתונים. אינטואיטיבית פיצ'ר זה הוא בעל השפעה על ההסתברות לביטול משום שקשה יותר לחזות מה יהיה בעוד חצי שנה מאשר מה שיהיה בעוד שבוע, על כן סביר שההזמנה של חצי שנה מראש לא תצא לפועל ומנגד סביר יותר שההזמנה שמבוצעת כשבוע מראש תמומש בסבירות גבוהה, מלבד ההיגיון שבדבר, ניתן לראות שגם הנתונים מציגים תמונה דומה.



הפיצ'ר השלישי שבחרנו הוא 'pay_now' ומתאר אם על הלקוח לשלם כבר במעמד ההזמנה או במועד אחר. הבחירה התבצעה בעיקר על סמך הנתונים ולא על האינטואיציה שלנו. חישוב הקורולציות העלה שלפיצ'ר 'charge_option' יש קורולציה גבוהה עם ההסתברות לביטול, בחרנו להנדס מעט את העמודה הזו כך שבמקום להציג את אחד מבין השלוש:

1. האם הלקוח ישלם מיד עם ההזמנה
2. הלקוח ישלם מאוחר יותר
3. מועד התשלום אינו ידוע

הגדרנו את העמודה מחדש כך שתתאר בצורה בוליאנית האם הלקוח ישלם מיד עם ההזמנה או לא וגילינו שכאשר הלקוח בוחר לשלם כבר במעמד ההזמנה:

ההסתברות לביטול יורדת מ 25% (תוחלת הביטולים כאשר לא ידוע דבר על סט האימון) ל 15%, כלומר יש מתאם גבוה בין הנתונים, על כן בחרנו להכניס גם את הפיצ'ר הזה לסט הפיצ'רים המצומצם שלנו.

