Cyber project

נושא החקר: הפתיחות ההכי פופולאריות וההכי מנצחות בשחמט.

מגיש: אביתר בוטין י'2

כשחקן שחמט (כבר 11 שנה) אני ושחמטאים אחרים יודעים שאחד החלקים הכי חשובים במשחק הוא הדביוט (הפתיחה) של המשחק. במשחק השחמט קיימות כ-1327 פתיחות שונות! על כל שחקן שחמט מותלת הבחירה הענקית הזאת של בחירת הפתיחה היחידה (או בחירה של האחדות) בה הוא ישחק מול יריביו הרבים ואיתה הוא ישאוף לנצח. ביום יום אנחנו יכולים לראות ששחקנים רבים משנים את הפתיחה שהם שיחקו זמן רב לפתיחה בה אין\ כמעט אין להם נסיון ולכן רובם מתחילים להפסיד ולרדת במד כושר שלהם (מד כושר הוא שיטה מקובלת לדירוג של שחקני שחמט, המבוססת על מודל סטטיסטי פשוט ויציב, והמנוהלת על ידי התאחדות השחמט העולמית (פיד"ה). לפי שיטה זו, לכל שחמטאי פעיל יש דירוג מספרי, המעודכן באופן קבוע לפי הניצחונות וההפסדים שלו במשחקים תחרותיים, באופן הלוקח בחשבון גם את רמתם של היריבים שמולם שיחק: ניצחון נגד שחקן חזק מקנה יותר נקודות מניצחון נגד שחקן חלש ממנו.) אז נשאלת השאלה: מדוע הם עושים זאת? ישנם כמה סיבות אבל העיקרית (לדעתי) היא מכיוון שלפני זמן רב הם עשו בחירה לא טובה של פתיחה אשר לא מתאימה להם, פתיחה אשר לא מספיק טובה או פופולארית. מסיבות אלא אני החלטתי לעזור לשחקנים חדשים ובעבודת חקר הזאת לבדוק על פי יותר מ 130000 משחקים מה היא הפתיחה הכי מנצחת ומה היא הפתיחה הכי פופולארית כדי שהם יוכלו ישר לעשות את הבחירה הנכונה ואז לא יצתרכו בעתיד להחליף אותה בגלל זה לרדת במד הכושר שלהם.

הנחות מרכזיות אישיות:

לדעתי הפתיחה הכי פופולארית היא Ruy Lopez Opening מכיוון שמהניסיון שלי שיחקתי אותה הכי הרבה פעמים גם בצבע הלבן וגם בצבע השחור.

לדעתי הפתיחה הכי מנצחת בלבן היא King's Pawn Opening: King's Knight, Elephant Gambit מכיוון ששיחקתי זמן רב את הפתיחה הזאת ואני יודע עד כמה היא יכולה להיות חזקה ובנוסף לכך אלוף העולם הנוכחי Magnus Carlsen משחק אותה ומראה בה היתרון גדול.

הפתיחה הכי מנצחת בשחור היא כנראה King’s Indian Defence מכיוון שהיא נותנת מצב בו הלבן צריך לשחק בלי טעויות כדי לנצח את המשחק.

תיאור מסד הנתונים

בעבודה אני השתמשתי בשתי דאטה סטים של משחקי שחמט. הדאטה סט הראשון הוא דאטה סט הבנוי מ48871 של משחקי שחמט אשר שוחקו על ידי מחשב מול מחשב; הדאטה סט בנוי מ - 19 עמודות:

(object)Game(str) – העמודה מייצגת על ידי מי שוחק המשחק, (object)White(str) – העמודה מייצגת את השחקן בכלים הלבנים,

(object)Black(str) – העמודה מייצגת את השחקן בכלים השחורים,

White Elo(int) - העמודה מייצגת את המד כושר של השחקן בלבן,

Black Elo(int) – העמודה מייצגת את המד כושר של השחקן בשחור,

White RD(float) – העמודה מייצגת את סטיית הדירוג של השחקן,

(float)Black RD - העמודה מייצגת את סטיית הדירוג של השחקן,

WhiteIsComp(bool) – העמודה מייצגת האם השחקן הוא מחשב,

BlackIsComp(bool) – העמודה מייצגת האם השחקן הוא מחשב,

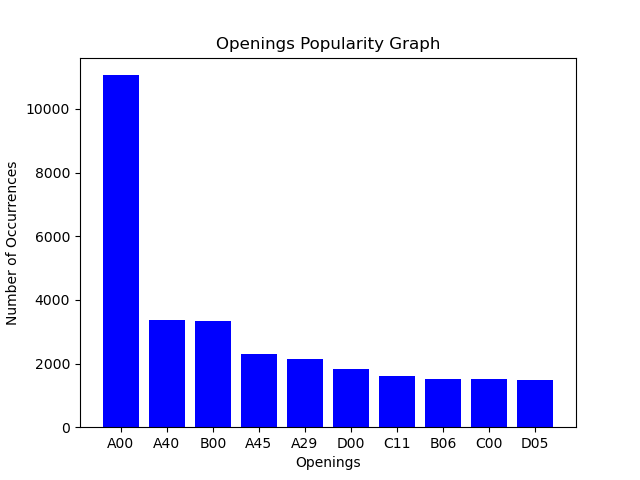
TimeControl(str) – העמודה מייצגת את כמות הזמן הניתנת למשחק בתחילתו וגם את התוספת לכל מהלך, Date(str) – העמודה מייצגת את התאריך בו המשחק שוחק, Time(str) – העמודה מייצגת את הזמן בו המשחק התחיל, White Clock(str) – העמודה מייצגת את הזמן של הלבן למשחק, Black Clock(str) - העמודה מייצגת את הזמן של הלבן למשחק, ECO(str) – העמודה מייצגת את הפתיחה אשר שוחקה(העמודה העיקרית בחקר זה),

PlyCount(int) – העמודה מייצגת את מספר המשחקים מול היריב הנוכחי בעבר, Result(str) – העמודה מייצגת את תוצאת המשחק, Commentaries(str) – העמודה מייצגת את תוצאה המשחק עם הערות נוספות, Move(str) – העמודה מייצגת את המהלכים של המשחק.

הדאטה סט השני מורכב מ81041 משחקים אשר שוחקו על ידי בן אדם מול מחשב; הדאטה סט מורכב מאותם עמודות כמו הדאטה סט הראשון.

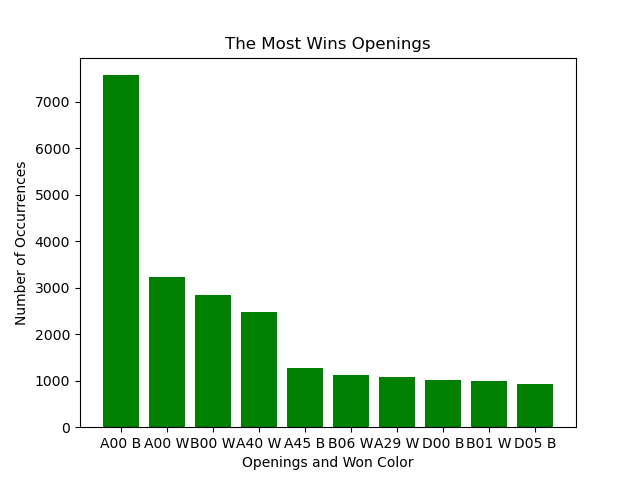
שתי הדאטה סטים נתונים בפורמט CSV לכן אנחנו בקלות יכולים לחקור אותם ולעשות את המניפולציות הרצויות לחקר שלנו בעזרת ספריית הנתונים PANDAS.

סיכום

לאחר עיבוד הנתונים הגענו למידע שרצינו, בהתחלה נעבור על הפופולאריות של הפתיחות: כ

אנחנו יכולים לראות שהפתיחה00A(King's Fianchetto Opening) היא הפתיחה הכי פופולארית והפופולאריות שלה עולה מעל ל11000(11061) שזה 8.5% מכל המשחקים ששוחקו! בנוסף לכך הפתיחה השנייה(40A Queen's Pawn Opening) קטנה ב 7,698(3363( מהפתיחה שזה רק 2% מהמשחקים מזה אנחנו יכולים להסיק כמה מסקנות הראשונה היא שטעיתי בהנחה הראשונית שלי כשחשבתי שהפתיחה 60C ( Ruy Lopez Opening) הכי פופולארית מסתבר שהפתיחה לא כל כך פופולארית והיא נמצאת במקום ה 138 עם בסך הכל 186 משחקים. המסקנה השנייה היא שהפתיחה 00A היא פתיחה מאוד פופולארית ועולה על אחרים בהרבה.

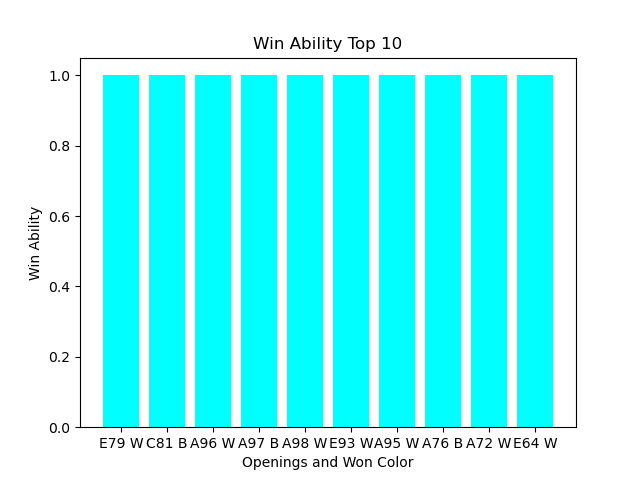
הפתיחות עם הכי הרבה ניצחונות:



אנחנו יכולים לראות שהפתיחה00A היא לא רק הכי פופולארית אל היא גם הפתיחה הכי מנצחת בשחור! במקום השני במספר הנצחונות נמצאת גם הפתיחה 00A רק בלבן מזה אנחנו יכולים להבין שלמרות שלשחק את הפתיחה הזאת יותר טוב בשחור כמות הפעמים בהם הפתיחה שוחקה (בשני הצבעים) מאפשרת לה להיות מעל כל שאר הפתיחות בשני הצבעים, ובמקום השלישי יש את הפתיחה 00B (King's Pawn Opening) בצבע הלבן, שהיא לדעתי הפתיחה הכי פופולארית למתחילים אבל היא בהחלט משוחקת על ידי רבי אמנים ואנשים עם מד כושר גבוה וכמו שאנחנו יכולים לראות משוחקת עם הצלחה גדולה.

גם מכאן אנחנו יכולים להסיק כמה מסקנות, הראשונה היא שכמעט צדקתי בהנחה בראשונית שלי לגבי הפתיחה המנצחת ביותר בצבע הלבן אך טעיתי לגבי הצבע השחור. המסקנה השנייה היא שהפתיחה 00A מנצחת יותר בצבע השחור.

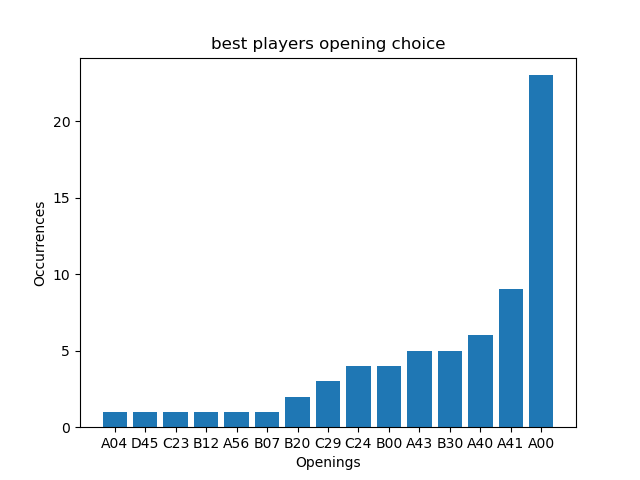
הפתיחות עם אחוז הניצחון הגדול ביותר:



מהגרף הזה אנו יכולים לראות שהפתיחות שראינו בשני הגרפים הראשונים לא מופיעות כאן אבל אנו יכולים להסביר זאת בכך שהפתיחות האלה פשוט משוחקות בהרבה פחות מהפתיחות המפורסמות יותר(אשר ראינו בשתי הגרפים הנל) לכן לדעתי צריך פחות להסתמך על הגרף הזה.

מצד שני אנו יכולים לראות שאחוז הנצחונות הגדול ביותר (100%) הוא אצל הפתיחות: 79E, 81C, 96A, 97A, 98A, 93E, 95A, 76A, 72A, 64E ו - 77E. עליהם אתם בהחלט יכולים לשים לב.

בחירת הפתיחות של השחקנים המצטיינים:



מהגרף הזה אנחנו יכולים לראות שהשחקנים עם אחוז הניצחונות הגדול ביותר בוחרים הכי הרבה את הפתיחה 00A שזה מוכיח שלא רק שחקנים פחות טובים משחקים את הפתיחה הזאת.

לסיכום המסקנות שאנחנו יכולים להסיק מהמסקנות הם שהפתיחה 00A היא הכי פופולארית ומנצחת הרבה לצבע השחור בנוסף לכך הפתיחה משוחקת גם על ידי שחקנים חזקים לכן זו היא פתיחה חזקה גם בשביל שחקנים ברמות הנמוכות וגם עבור שחקנים ברמות הגבוהות לכן אני מציע ואני בעצמי אתחיל לשחק את 00A בצבע השחור בנוסף הייתי מציע לכל שחקן לדעת את הפתיחה גם בצבע הלבן מכיוון שהיא משוחקת על ידי רבים בשתי הצבעים. לגבי הצבע הלבן אנחנו ראינו שהפתיחה הכי מנצחת היא 00B והיא גם פתיחה די פשוטה בשביל ההתחלה לכן אני מציע לכל שחקן לשים עליה לב.

רפלקציה

בחירת נושא:

לפני תחילת העבודה היה לי רעיון שונה למחקר אשר היה קשור לשפעת הספרדית ולמגפת הכולרה אך אחרי חיפושים רבים מצאתי את הדאטה סטים של משחקי השחמט והחלטתי לשנות את נושא המחקר.

קשיים עיקריים:

במהלך העבודה הקושי העיקרי שחוויתי היה בחלק השלישי של העבודה (בתהליך כתיבת הקוד) מכיוון שבעבר לא באמת נפתחתי לנושא הביג דאטה(חוץ מהבית ספר) ולכן לא הייתי חופשי בידע ובשימוש בספריות הקשורות לנושא אך החבר הכי טוב שלי (אדון גוגל) עזר לי וכך התגברתי על הקושי.

החלק בו הכי נהניתי:

החלק שבו הכי נהניתי היה גם החלק השלישי מכיוון שנגליתי ליכולות הגדולות והיחסית פשוטות למימוש של עבודה וחקר של מקור מידע גדול מאוד.

החלק בו הכי לא נהניתי:

כתיבת רפלקציה... 😊.

לסיכום, נהניתי מתהליך ותוצאות המחקר.

קישור למאגר הנתונים:

<https://data.world/thule179/2016-internet-chess-games/workspace/file?filename=ficsgames_analysis.R>

גרפים

