

## דוח מסכם- ניתוח נתוני דירוג קפה arabica

### 1. מבוא- הצגה של מאגר הנתונים:

עבדתנו מבוססת על ניתוח נתונים ציבורי הכלול 1,318 טעימות קפה מסווג ערביתה, אשר נבדקו על ידי טועמים מוסמכים בהתאם לפרטוקול הרשמי של SCA- איגוד הקפה המובחר. הדגימות נאספו בין השנים 2004 ל-2018 וכל שורה מייצגת טעימה אחת. כל דגימה כוללת ציונים במספר מאפיינים חשובים (סנסוריים) כמו טעם, גוף, חומציות ואייזון, לצד פרטיטים טכניים הנוגעים למקור הקפה - מדינת הגידול, זן, גובה הגידול, שיטת העיבוד ועוד. המאפיינים החשובים מודרגים על פי פרטוקול של איגוד הקפה בלבד המופיע 'Cupper Points' שמדרג לפי התרשומות כוללת.

מערך הנתונים זמין דרך [github](https://github.com/jdbc/coffee-quality-database/tree/master).

### 2. ניתוח מקדים של הדadata A:

בשלב זה ביצענו חלוקה גסה של הנתונים לנתונים סנסוריים ונתונים טכניים. בנוסף, בחנו את התפלגות הנתונים וניסינו לאפיין דפוסי התפלגות שיסיעו לנו בשאלת המחקר.

#### 1.1. ביקיון , סידור ונירמול הדadata:

בשלב הראשון בחנו את מבנה הנתונים והאם יש בו בעיות שימושיות. הושרו עמודות עם שיעור גובה מאוד של ערכי חסר וכן עמודות שכלי הערכים בהן זהים או לא רלוונטיים לשאלת המחקר (למשל מהיה בעליים, פרטי משקל ועמודות מזהות קללאים). ערכי הגובה הומרו ליחידות אחידות (מטרים), האותיות באנגלית שונו לאותיות קטנות בלבד ופורמתי התאריכים סודרו. המטרה: לוודא שאין כפיליות בעמודות, המשתנים אחידים , אינפורטטיביים ומוכנים להמשך עבודה על בסיס נתונים תקפים ומובנים.

#### 2. בוחנת התפלגות של משתנים מרכזיים בDATA:

##### א. משתנה סנסורי- ציון:

1. ציון אובייקטיבים- הציונים מודרגים לפי פרטוקול קבוע במהלך הטעימה עברו מאפיינים סנסוריים כמו Acidity, Flavor , Aftertaste , Aroma , Balance ועוד (ישנם 9 מאפיינים).

2. ציון סובייקטיבי כולל- Cupper Point, ניתן לאחר דירוג שאר המאפיינים הסנסוריים ומבטא את התרשומות הטועם מהקפה בכללותן.

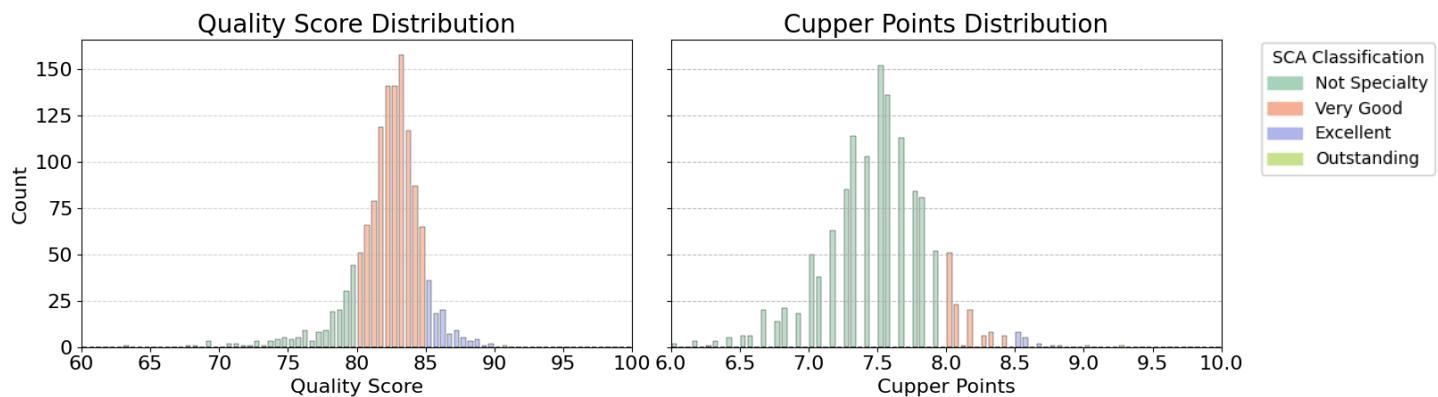
שני סוגים הציונים נעים בטוחה של 6 עד 10 ובקפיצות של 0.25. ציון של 10 מייצג איכות מקסימלית, בעוד שציון של 6 הוא הסף התיכון שבו הקפה עדין נחسب ראוי להערכתה לפי איגוד הקפה המובחר.

בהתאם לסכימת כלל הציונים הסנסוריים, מתקיים Quality Score וציוני נוע בין 100-60. לציונים הכלולים קיימת סקללה של איגוד הקפה המובחר על מנת לדרג אותם באופן איכותי.

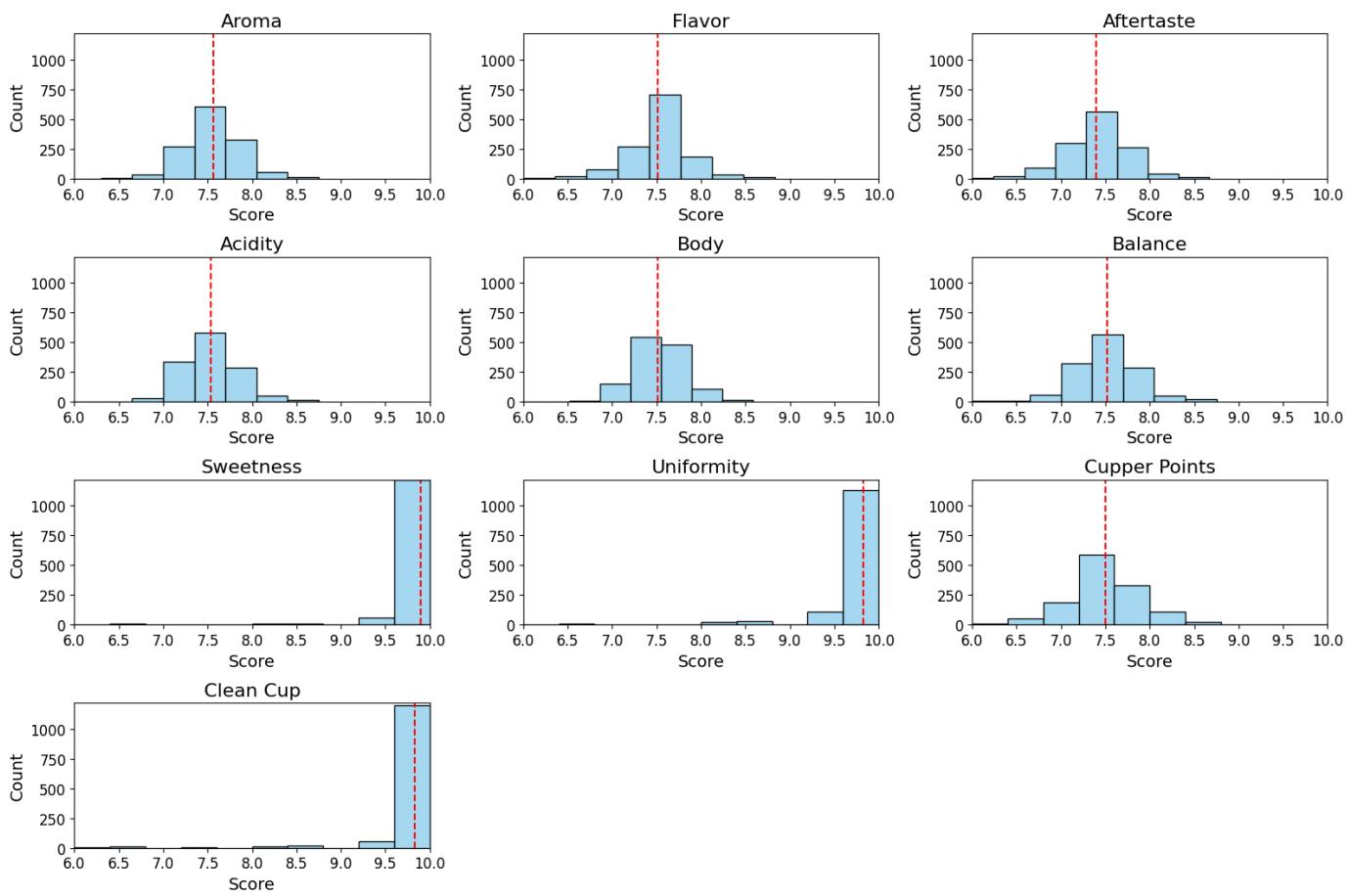
א. מהסתכלות על התפלגות הציונים הכלולים בתרשימים מספר 1 ניתן להבחן שרוב (77.7%) הציונים שניתנו היו בסולם האיכותי-'Very Good'. לעומת זאת, כאשר בוחנים את התפלגות ציוני ה-Cupper Points, עלולה כי רוב (89.5%) הדגימות דורגו כ-'Not Specialty', כלומר מתחת לרף האיכות של איגוד הקפה המובחר. ניתן זה עשוי להיות כי קיימים גורם נוספת לטעימה הסנסורית המשפיע על הציון הסובייקטיבי.

ב. מתרשים מספר 2 ניתן לראות כי מרבית המאפיינים (כגון Aroma, Flavor, Acidity) נעים סביב ממוצע של 7.5-8 ובעלתי פיזור דומה יחסית לציון הסובייקטיבי. בעוד מאפיינים כמו Sweetness, Uniformity,

Clean Cup מציגים ציון כמעט מושלם של 10 לרוב הדגימות - דבר המעיד על חוסר שונות במדדים אלה, מה שעשוי להפחית את תרומתם להסביר ההבדלים בין דגימות.



**תרשים מס' 1: התפלגות הציונים במרק הנטוינטס .** בצד השמאלי מוצגת התפלגות הציונים הכלולים (Score Quality), ובצד הימני התפלגות ציון Copper Points, תוך שימוש בקפיצות מותאמות: 0.5 ל- Score (Score), ו-0.05 ל-Copper Points. חלוקת הצבעים נעשתה על פי חלוקה איקוונית של איגוד הקפה המובחר.



**תרשים 2: התפלגות מאפיינים סנסוריים.** בתרשימים מוצגת היסטוגרמה של כלל הציון הסנסוריים. כאשר הקוו האדום בכל גרף מצין את הממוצע של אותו מאפיין.

#### ב. מאפיינים טכניים:

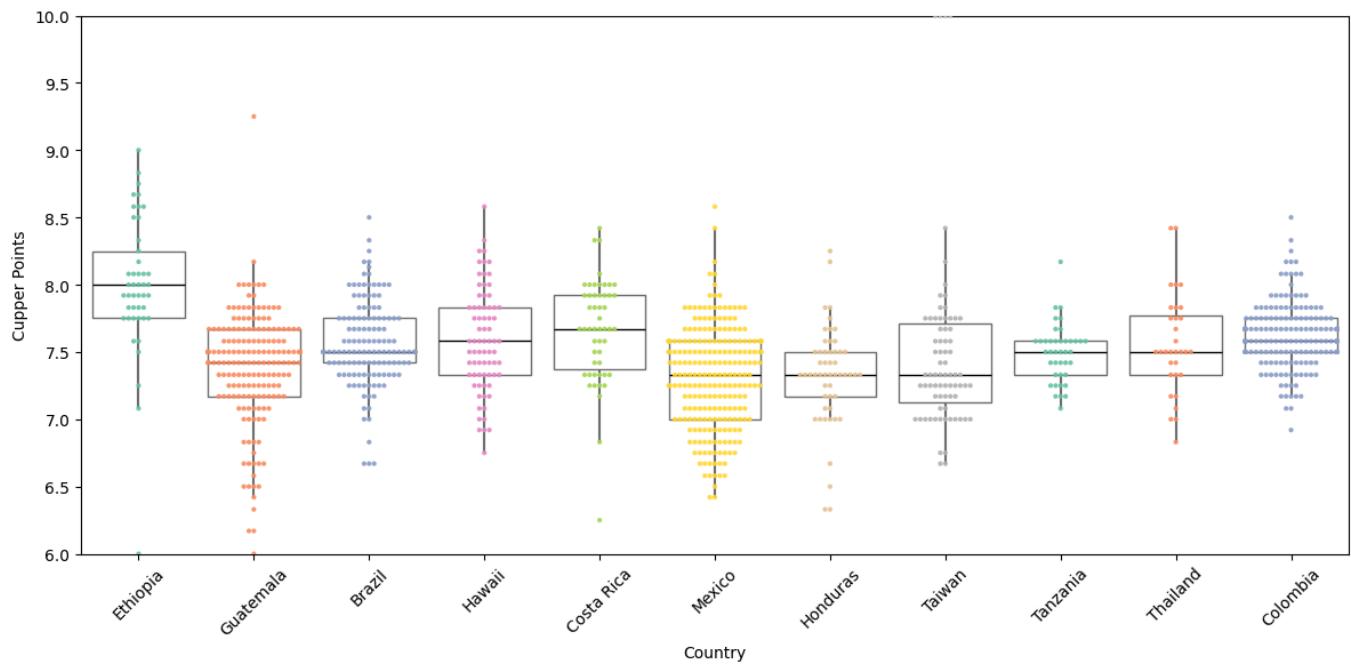
המאפיינים הטכניים של פולי הקפה כוללים משתנים כמו רציפים כדוגם גובה גידול (800–2500 מטר), אחוז לחות (10%-12.5%), וכן משתנים שימושיים כדוגם זן הקפה (זן ערביתה שונים), אזור גידול, שיטת עיבוד ומדינה המקור.

פרמטרים של אחוז הלחות בקפה וצבע הקפה הם פרמטרים שנמדדו במעבדה ולא היו גלויים לשופטים. בפרויקט זה בחרנו להתמקד על השפעתם של הפרמטרים – מדינת המקור, מאפיינים סנסוריים ושיטת העבודה.

#### מדינת המקור

פולי הקפה גודלו ב-36 מדינות שונות. לאחר שהחלק מהמדינות מספור הדגימות קטן מאוד, סיננו את הנתונים כך שיישארו לניטוח רק מדינות עם יותר מ-30 דגימות. סוף זה מפחית את ההשפעה של רוש סטטיסטי ומבטיח השווואה יציבה ומהינה. לאחר הסינון נותרנו עם 11 מדינות, בהן מספר התוצאות אפשרן ניתוח משמעותי יותר. בתרשימים ניתן לראות כי קיימת שונות ניכרת בין המדינות מבחינה רמות הציון הסובייקטיבי (Copper Points) וגם ברמת הפיזור הפנימי. על ידי כך ניתן לזהות דפוסי פיזור שונים:

- I. פיזור אחד ורחב - כגון מקסיקו  
 II. ריבוי ערכיהם חריגים - כגון הונדורס.  
 III. פיזור גובה עקבי - כגון אתיופיה  
 בנוסף, רוחב התיבת הגדולה של מקסיקו וטאיוואן מצביע על שונות פנימית בגובה במינוח.  
 גיון התוצאות בין המדינות מעיד על שונות ברמת האיכות והאחדות של הקפה הנמדד.



תרשים מס' 3: התפלגות ציוני הטעימה (Cupper Points) ופיזור הדגימות לפי מדינת המקור. בתרשימים מוצגים חציון, טוח בין רביעוני וערכים חריגים.

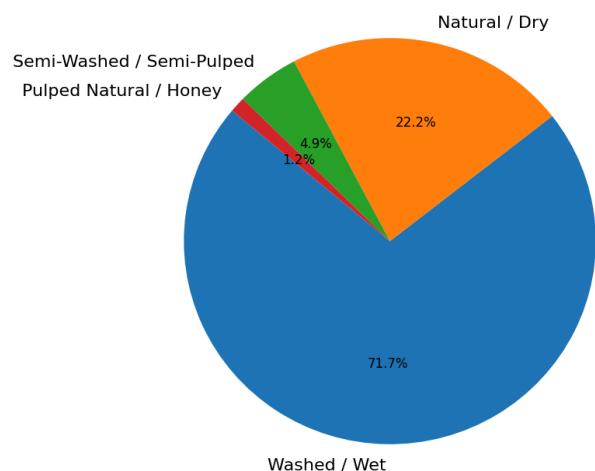
#### שיטת העבודה

במערך הנתונים מופיעות ארבע שיטות עבודה שונות לפולוי הקפה תהליך חשוב המשפיע הן על מאפייני הטעם והן על איכות הקפה. למאפייני שיטת העבודה עשויה להיות השפעה על ציוני הטעימה (Cupper Points), שכן עבודה שונה עשוי להציג או לטשטש תכונות מסוימות של הקפה.

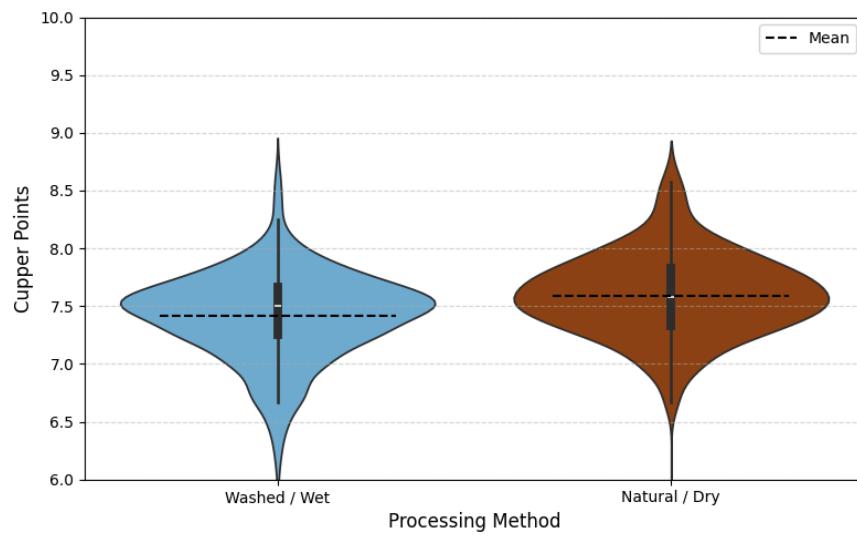
- I. מתרשים מס' 4 ניתן להבחין כי השיטה הנפוצה ביותר היא Wet / Washed, אחרת נמצאת / Natural Dry והשאר מתפלגות בין שיטות פחות שכיחות. בחרנו להתמקד בשתי השיטות המרכזיות, שכן מהוות את רוב הדאטה ובכך מאפשרות ניתוח מובהק, יציב ואמין יותר. בנוסף, סינון זה מסייע לצמצם את ההשפעה של מקיריות או שונות גובהה שעוללה לנובע ממספר תכיפות קטן בשיטות הפחות שכיחות.  
 II. מתרשים 5 ניתן להבחין בתפלגות השיטות ביחס לציון הסובייקטיבי - שיטת Wet / Washed מציגה פיזור רחב יותר של ציוני Cupper Points, כולל ערכים חריגים נזוכים. לעומת זאת, שיטת Dry / Natural מציגה פיזור צפוף ומרוכז יותר בטוחה הציון הגבוהים.

למרות שהמוצع בשיטה הטבעית מעט גבוה יותר, פער גודל המדגם לא מאפשר הסחת מסקנות חד משמעיות, אך כן מצביעים על מגמה ראשונית מעניינת.

**תרשים מס' 4: התפלגות שיטות העיבוד של פול' הקפה בمعרך הנתונים.**



**תרשים מס' 5: התפלגות ציוני Copper Points לפי שיטת עיבוד.** בתרשימים מסומנים חץין, ממוצע, רביעון העליון והתחתון.



### 3. שאלת המחקר

לאחר ביצוע ניתוח ראשוני (EDA) להבנת התפלגות הנתונים,>Zיהוי דפוסים וחריגות, החלטנו להתמקד בשאלת: האם ניתן לבא את ציוני-h-dots Cupper Points בקפה ערבייה על בסיס המאפיינים הסנзорיים בלבד (כגון אroma, טעם, אופטרטיסט, חומציות, גוף ואיזוז) או שמא קיימים גורמים נוספים, כמו שיטת העיבוד ומדינת המקור, המשפיעים על הציון הסובייקטיבי ומרמזים על קיומו של אפקט מוניטין סמי?

### 4. ניתוח מרכזי

הניתוח המרכזי בוצע בשלושה שלבים עיקריים. בשלב הראשון, ניתחנו את האינטראקציה והתרומה של כל מאפיין בנפרד (פרמטרים סנзорיים, מדינת מקור ושיטות עיבוד) לצינוי הטיעמה הסובייקטיבים. בשלבים הבאים השתמשנו ב מבחני רגרסיה על מנת לכמת את הקשר ולבוחן את תרומתו היחסית של כל פרמטר כחלק ממודל רב משתני.

#### 4.1. תרומת כל מאפיין לציון הסובייקטיבי:

בניתוח המקדים הבחנו בהתפלגויות שונות עבור משתנים בעלי אותה קטגוריה ביחס לציון הסובייקטיבי. על מנת לבחון את זה ביצעו אנליזה התחלית בין משתנים שונים המשתייכים לאוֹתָה קטגוריה (מ בין הקטגוריות- מאפיין סנסורי, מדינה ושיטת עיבוד). המטרה הייתה לבחון האם יש "תרומה" שונה של המשתנה לציון הסובייקטיבי.

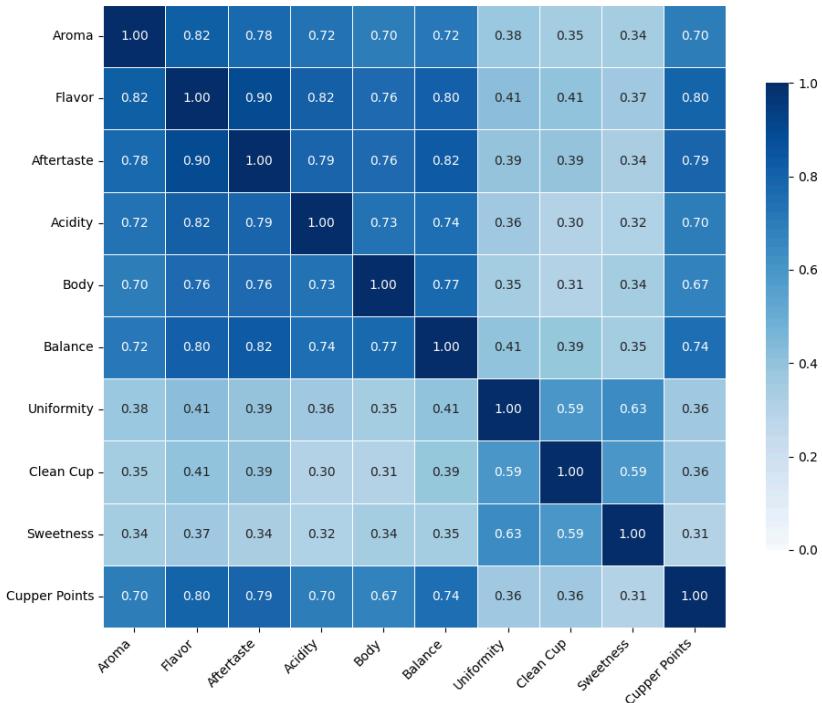
#### א. קורלציה של המאפיינים הסנסוריים לציון הסובייקטיבי:

תרשים מס' 6 ניתן להסיק אילו פרמטרים סנסוריים עשויים להיות החשובים ביותר בתחום של **Copper Points**. ניתן לבחון כי יש שישה משתנים בעלי ערך מתאים גובהו בהתאם הגובה ביותר הוא 0.80. למאפיין **Flavor**.

שלושה משתנים נמצאו עם מקדם מתאם נמוך- **Sweetness**, **Clean Cup**, **Uniformity**. עובדה זו מבוססת את ההנחה הראשונית שלנו יתכן ואלו פרמטרים שפחות משפיעים על הציון הסובייקטיבי. לעומת זאת מנגנון הראשון שלנו לתרשים מס' 2- יתכן כי מאפיינים אלה מהווים רעש.

#### **תרשים מס' 6 : מקדמי מתאם בין המאפיינים הסנסוריים לבין הציון הסובייקטיבי (Copper Points).**

הערכים מוצגים כערך מתאם פירסום, כאשר צבע כהה מייצג מתאם חזק יותר בין המשתנים.

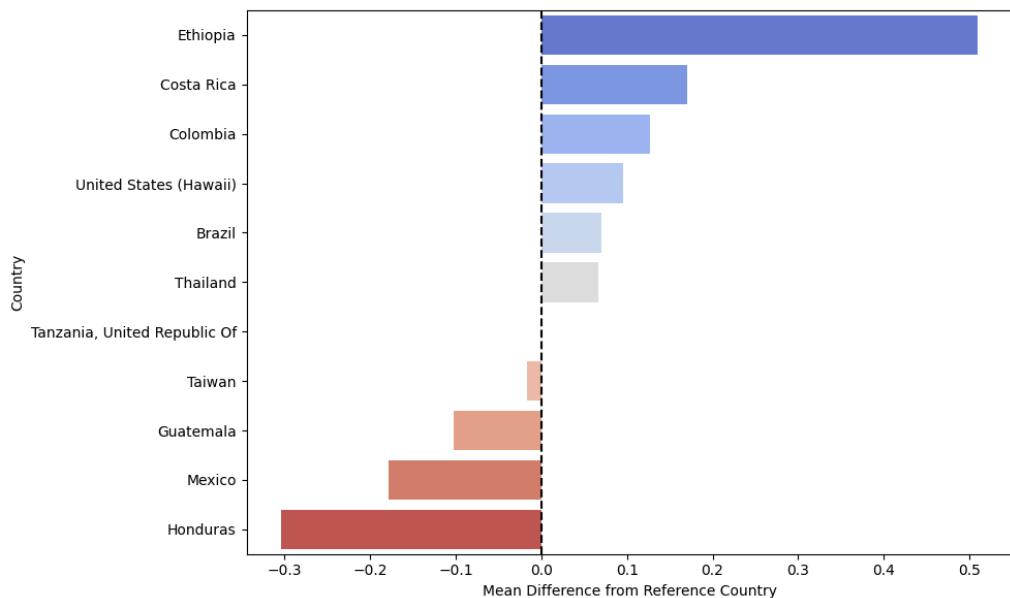


#### ב. התרומה של ידעת מדינת המדינ

#### לציון הסובייקטיבי:

תרשים מס' 7 מוצגים "התרומות" של כל מדינה לציון **Copper Points**. ממוצע **Copper Points** של כל מדינה נבחן ביחס למדינות ייחוס. עמודות כחולות מייצגות מדינות עם ציונים גבוהים יותר מה ממוצע, ועמודות אדומות מייצגות מדינות עם ציונים נמוכים יותר. לצורך ההשוואה נבחרה מדינת ייחוס ניטרלית, טנזניה. היא המדינה עם ממוצע ציוני **Copper Points** הקרוב ביותר ממוצע הכללי של כל המדינות שנדרשו לאחר הסיכון. בחירה במדינה ניטרלית נעדרה לאור העובדה שהיא השפעה יחסית ולא מוחלטת.

אטיפיה, קויטה ריקה וקולומביה הנוטות מפער חיובי משמעותי ותורמות לעלייה בציון הסובייקטיבי. לעומת זאת, מדינות כמו הונדורס, מקסיקו וגואטמלה תרמו באופן שלילי וגרמו לציונים ממוצעים נמוכים יותר.



**תרשים מס' 7: השפעת מוניטין מדינית המקור על הציון הסובייקטיבי - השפעה חיובית שלילית או נייטרלית. מוצגים פער ממוצע ביחס למדינת ייחוס (טנזניה), שנבחרה לפי קרבה סטטיסטית לממוצע הכללי.**

**ג. התרומה של דיאית העיבוד לדירוג הסובייקטיבי:**

בתרשים מס' 8 מוצג הפער הממוצע בין שיטות העיבוד עבור כל מאפיין סנסורי. הפער הגבוה ביותר נמצא בציון Cupper Points, מה ש מעיד שדיאית העיבוד "מוסיפה" יותר נקודות לציון הסובייקטיבי לעומת מאפיינים אחרים. גם Metrics פערים愕ולטימ, מה שעשו להיעד על השפעה עקיפה או חפיפה בין המדרדים.

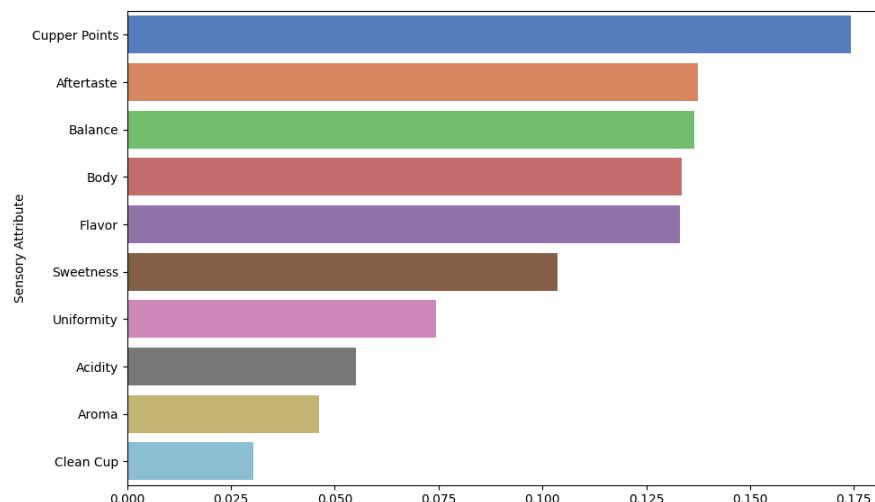
כדי לבדוק זאת באופן מדויק יותר, ביצענו מבחן פרטוטזיות עם 100,000 חרוזות Über Quality Cupper Points Score (טבלה 1). בשני המקרים התקבלה מובהקות סטטיסטית, אך ב-*Cupper Points Score* הייתה חזקה יותר - מה ש מעיד על הבדל עקבי בין שיטות העיבוד. מנגד, הפער הממוצע היה גבוה יותר דווקא ב-*Quality Score*, דבר שעשוי להיעד על השפעה רחבה יותר של שיטת העיבוד גם על רכיבים שאינם סובייקטיביים בלבד.

**תרשים מס' 8: פער ממוצע**

**בציוני מאפיינים סנסוריים בין**

**שיטות עיבוד - כמה "נקודות"**

**מוסיפה לשיטת העיבוד לכל מאפיין.**



סוג הציון	P-value	פער ממוצעים שנכפה- בערך מוחלט
Cupon Points - ציון סובייקטיבי	0.00000	0.174
Quality Score - ציון כולל	0.00376	0.606

טבלה מס' 1: תוצאות מבחן פרמטריות (100,000 חוות) לבחינת השפעת שיטת העבודה על Cupper Points ו-Quality Score.

#### 4.2. בדיקת רגסיה לינארית:

לאחר ניתוח פנימי של כל קטגוריה, ביצענו השוואה בין הקטגוריות השונות באמצעות מודל רגסיה לינארית מרובת משתנים מסוג LASSO, שונען לחזות את ציון ה-*Cupper Points*. מודל זה מניח קשר לינארית בין המשתנים ומבצע רגוליזציה מסוג L1 שמטרתה לצמצם מורכבות יתר, להפחית את השפעתם של מקדים חשובים ולאפס משתנים שאינם תורמים לחיזוי. בנוסף, נעשה שימוש ב-*Cross Validation* לצורך קביעת עצמת הענישה (למבדה) באופן אופטימלי וכן מתבל מודול מדויק יותר שאינו מותאם יתר לננתוני האימון. כדי לבנות מודל שחזק טוב גם על נתונים חדשים, הנתונים פוצלו ל-80% עבור סט אימון ו-20% עבור סט בדיקה באמצעות חלוקה אקראיית.

לפי טבלה מס' 2, התוצאות שהתקבלו ממחישות את חשיבותם של המשתנים הסנסוריים כגורם מנבأ עיקרי. מודל 1 השיג מקדם הסבר של 0.739, תוצאה המעידת על יכולת חיזוי גבוהה יחסית בהתבסס על משתנים אלה בלבד.

הוספת המשתנים של מדינת המקור ושיטת העבודה (מודלים 2,3) הביאה לשיפור זניח בביוציע המודל והעלו אותו מעט. פערים קטנים אלו מצביעים על כך ש מרבית המידע הדרוש לחיזוי כבר כולל במאפיינים הסנסוריים עצםם וכן יש צורך עבור לשיטה לא לינארית.

בנוסף לביוציע החיזוי, נצפו גם ערכי למבדה שונים בכל אחד מהמודלים, אשר משקפים את עצמת הענישה שביצעו המודל על מקדים הרגסיה. מודל 1 קיבל ערך למבדה של 0.00762. מודל 2 הביא לירידה קלה בערך זה ואילו מודל 3 הוביל לערך נמוך מאוד.

ירידה בעוצמת הענישה מצביעה על כך שככל שנוספו משתנים נוספים למודל נדרש פחות כיווץ אגרסיבי של המקדים. עם זאת, הירידה הדרגתית וערך הלמבדה הנמוך מאוד במודל 3 משקפת מצב שבו יש מידע מסוים נוספת אך הוא חופף מאוד למה שכבר קיים וכן המודל מעוניין בצורה מתונה יותר אך עדין מדכו מושתנים שאינם תורמים תרומה "יחודית לחיזוי".

ניתן להסביר את הפער בין התוצאות באמצעות אופן הפעולה של מודל LASSO. מודל זה פחות רגש לרעש ולכן אינו מתייחס אליו כגורם משמעותי ומאפס אותו. כאשר קיימים משתנים בעלי מידע חופף, המודל יעדיף לשמור רק את אחד מהם, בדרך כלל את המשתנה המסביר בצורה ישירה יותר את השונות בציון ובכך ימנע כפילות ויבנה מודל חסכוני, פשוט ועמיד לרעש, תוך שימוש במינימום משתנים.

$R^2$	למבדה	מספר המשתנים שהורצו	המשתנים שנבדקו	מו' המודל
0.739	0.00762	9	סנסורים	1
0.741	0.00618	20	סנסורים + מדינה	2
0.742	0.00056	22	סנסורים + מדינה + שיטת עיבוד	3

טבלה מספר 2: **תוצאות ריצת רגסיה לינארית (LASSO)** עם קבוצות משתנים שונות. המודלים נבנו על בסיס משתנים סנסורים בלבד, בשילוב מדינת המקור ובשילוב גם שיטת העיבוד (עבור שתי השיטות הנפוצות ביותר). הטבלה מציגה את ערך Lambda שנבחר בקורס וlidzha ואת מקדם ההסביר  $R^2$ .

4.3. **מודל לחיזוי ציון ה-Cupper Points:** בוחנת אינטראקציות בין המשתנים השונים בשלב זה בוחנו כיצד שילוב משתנים מקטגוריות שונות (מאפיינים סנסורים, מדינת המקור ושיטת העיבוד) משפיעים על דיקוק החיזוי. ככלומר, לבדוק האם קיימות אינטראקציות בין משתנים משודדים שונים. לשם כך נעשו שימוש במודל Random Forest, שמסוגל לזהות קשרים לא לינאריים וללכוד אינטראקציות מורכבות. לפני בניית המודל, הסרנו משתנים סנסורים שעולאים להווות רעש או חיפוי, כמו Chopping, Uniformity ו-Cup Sweetness. בהתאם למסקנות מהשלב הראשון של ניתוחה. עם זאת, בחרנו להשאיר את כל מדינות המקור (מדינות מעל 30 צפיפות) כדי לבחון השפעות מגוונות ולהזות פערים בין מדינות.

טבלה מספר 3: כאשר מתבצע רק מודל 1 מתקיים  $R^2 = 0.701$ , בקבוצת הבדיקה. הוספת מדינת המקור תורמת לשיפור קטן וזניח. השיפור המשמעותי נצפה במודל 3 בו הערך מזנק ל-  $-0.743 = R^2$ . ממצא זה מרמז על אינטראקציה בין שיטת העיבוד לבין משתנים אחרים המשפיעה על הציון הסובייקטיבי. כדי להימנע מ-Oversampling, כוננו את המודל לפתרמים שמרניים יחסית.

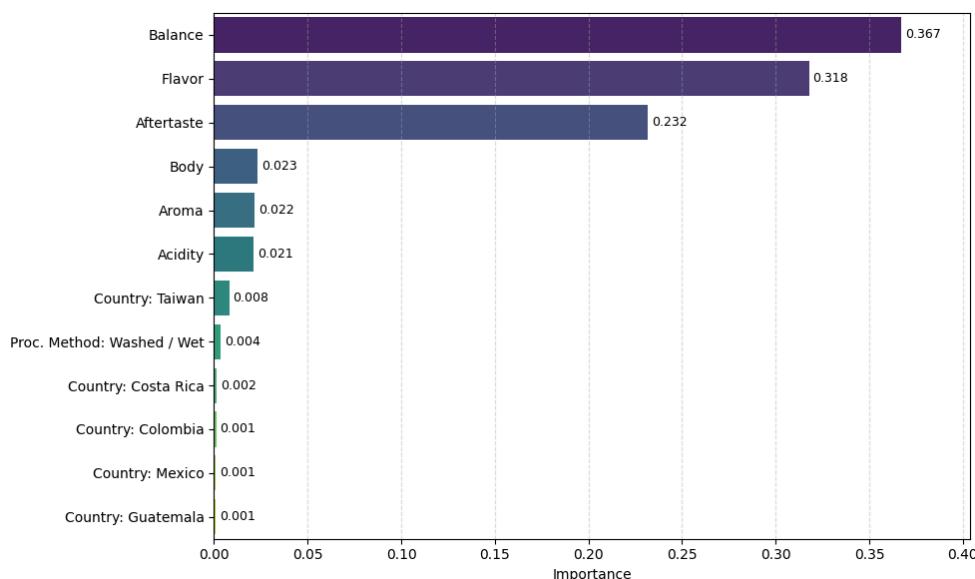
בהתואה למודל הלינארי (LASSO) ניתן להבחן בשיפור ממשמעותי בדיקוק החיזוי בעת שימוש במודל לא לינארי. אמן ערך  $R^2$  בקבוצת הבדיקה עולה רק מ-0.742 ל-0.743, אך יכולת הכללת האינטראקציות הלא לינאריות ערך  $-R^2$  הגבוה יותר בקבוצת האימון מצביעים על התאמת טובה יותר של Random Forest למורכבות הנתונים.

בשני המודלים הראשונים ערכיו  $-R^2$  באימון ובמבחן דומים יחסית, מה שמעיד על איזון טוב בין למידה להכללה. לעומת זאת, במודל ה-3 נוסף פער ממשמעותי בין האימון למבחן, שמרמז על תחילת Overfitting. יתר על כן מהוועפת משתנים רבים מדי ביחס לגודל הדגימה. על כן נתייחס לתוצאות המבחן ( $R^2$ ) כמדד העיקרי להערכת ביצועי המודל, שכן הן מייצגות את יכולתו הכלכלית לחזות נתונים חדשים.

בתרשים מספר 9 ניתן לראות, שלושת המאפיינים הסנסורים Balance, Flavor ו-Aftertaste (הם בעלי החשיבות הגבוהה ביותר). יחד הם תורמים למעלה מ-90% מהחישיבות המוסברת במודל. משתנים טכנים כמו מדינת המקור ושיטת העיבוד כמעט ואינם תורמים ישירות לניבוי, דבר שמחזק את ההנחה כי השפעתם על הציון הסובייקטיבי נעשית בעקיפין דרך השפעה על מדדים סנסורים אחרים.

TEST R <sup>2</sup>	TRAIN R <sup>2</sup>	מספר המשתנים שהוריצו	המשתנים שנבדקו	מו' המודל
0.701	0.624	6	סנסורים	1
0.702	0.628	17	סנסורים + מדינה	2
0.743	0.808	19	סנסורים + מדינה + שיטת עיבוד	3

טבלה מס' 3: **תוצאות ריצת מודל Random Forest עם קבוצות משתנים שונות.** לצורך הפחיתת רעש ואובר-פיטינג, המודל הוגבל לעומק מקסימלי של 6 (max\_depth=6) ולמספר מינימלי של 5 תצפיות בכל עלה (min\_samples\_leaf=5), כך שנשמר איזון בין ביצועים לבין יכולת הכללה.



תרשים מס' 9: **דירוג החשיבות של המשתנים במודל לחיזוי ציון Cupper Points.** בתרשימים זה מוצגים לפי סדר חשיבות כלל המשתנים עליהם התבסס מודל החיזוי Random Forest אשר חזה את דירוג הציון הסובייקטיבי באמצעות משתנים סנסורים+מדינה+שיטת עיבוד.

## 5. מסקנות:

מכלול המודלים והניתוחים שביצענו עולה כי ניתן להסביר חלק ניכר מהשינויים בציון הסובייקטיבי Cupper Points באמצעות מאפיינים סנסוריים בלבד, תוך שמירה על איזון בין דיקוק לבין יכולת הכללה. עם זאת, הוסף משתנים טכניים, בעיקר מדינת המקור ושיטת העיבוד, הביאו לשיפור מתון בלבד במידת ההסביר במודל הלינארי,

אך לשיפור משמעותו יותר במודל הלא ליניארי (Random Forest). מצא זה מרמז על קיומן של אינטראקציות או השפעות לא ליניאריות בין תחושת הטעם לבין מידע טכני, יתכן שאף השפעות עיקריות או אפקטים של "מונייטין סמי".

במהלך ניתוח הנתונים בחנו את התפלגות Cupper Points עבור כל קטגוריה (מדינה ושיטת עיבוד) בנפרד ויזיהינו הבדלים מובהקים. עם זאת, יתכן כי השפעות אלו חופפות לאפיון הסנסורי של הדגימות. למשל, יתכן שהמדינה או שיטת העבודה משפיעות על פרופיל הטעם ודרך על הציון אך לא ניתן להסיק על קשר סיבתי ישיר בין המידע הטכני לבין הציון שניתן בפועל.

לסיכום, צוינו הקפה ניתנים במידה רבה על סמך פרופיל חושי עקבי, אך יתכן כי קיימות גם השפעות עיקריות של מידע טכני שהזיג לשופט. בכרך, מצוי הפרויקט מצביים על אפשרות לקיום אפקט של מונייטין סמי בתחום הערכה.

בהמשך למחקר הנוכחי, ניתן להעמיק את הניתוח בשתי דרכי עיקריות: מצד אחד, לבחון את תרומותם של מאפיינים נוספים הקיימים בדעתה כמו גובה, לחות, יבואן, שנת קציר ועוד להערכת הטעם. מצד שני, ניתן לבחון האם קיימת עקבות בין הדירוג הסובייקטיבי לבין הדירוג הקטגוריאלי הקפה לפי סקלת האיכות של איגוד הקפה (Specialty / Very Good / Excellent / Outstanding). ניתוח זה עשוי לשפוך אור על הערות בין ציון מסווני לדירוג איכות ולחשוף מקרים שבהם פרופיל הטעם אינו תואם את סיווג הקטגוריה מה שעשו להציג על הטיות סובייקטיביות או סטנדרטים לא אחידים בתחום הערכה.

לינק לקובץ -

<https://colab.research.google.com/drive/16CdByTq605Ss6iPICJqUsbbr160Osslw?usp=sharing>